On the rigidity of amorphous solids. Price fluctuations, Conventions and Microstructure of Financial Markets

Matthieu Wyart

Thèse presentée pour obtenir le titre de

Docteur de l'École Polytechnique

specialité: Physique

par

Matthieu WYART

Sur la Rigidité des Solides Amorphes

&

Fluctuations des Prix, Conventions et Microstructure des Marchés Financiers

Soutenance prévue le 24 Novembre 2005 devant le jury compose de

MM. Jean-Louis BARRAT	Rapporteur
Alan KIRMAN	Rapporteur
Eric CLEMENT	
Marc MÉZARD	
André ORLÉAN	
Jean-Philippe BOUCHAUD	

Note Explicative

Cette thèse est écrite en anglais et contient deux parties distinctes. La première partie s'intitule ``Sur la Rigidité des Solides Amorphes'' et la seconde ``Fluctuations des prix, Conventions et Microstructure des marches Financiers''. Chaque partie est succédée par un article dont la thématique est voisine.

Sur la Rigidité des Solides Amorphes:

On comprend mal les propriétés microscopiques des solides amorphes, comme le transport, la propagation des forces ou la nature de leur rigidité mécanique. Ces questions semblent liées à la présence d'un excès de modes vibratoires à basse fréquence, le ``pic boson''. On explique la nature de ces modes dans les systèmes répulsifs à courte portée. On argumente que cette description s'applique aussi aux milieux granulaires, à la silice, et aux verres colloïdaux.

Fluctuations des prix, Conventions et Microstructure des marches Financiers:

Les fluctuations des cours de la bourse ont des propriétés étonnantes. La volatilité (l'amplitude de ces fluctuations) est environ un ordre de grandeur plus grand que les prédictions de la théorie des marches efficients, et est corrèlee sur des échelles de temps très longs. Les agents sur réagissent aux informations. On montre que ces propriétés apparaissent lorsque les agents agissent en fonction de leur expérience et du passé du marché. On étudie aussi la microstructure des marchés, qui régulent les échanges aux temps courts. On explique pourquoi le prix est diffusif bien que les ordres marchés (les chocs subis par les prix) soient très corrélés. On évalue la fourchette des prix par des arguments de symétrie.

Part I: On the rigidity of amorphous solids

December 6, 2005

Contents

Abstract

We poorly understand the properties of amorphous systems at small length scales, where a continuous elastic description breaks down. This is apparent when one considers their vibrational and transport properties, or the way forces propagate in these solids. Little is known about the microscopic cause of their rigidity. Recently it has been observed numerically that an assembly of elastic particles has a critical behavior near the jamming threshold where the pressure vanishes. At the transition such a system does not behave as a continuous medium at any length scales. When this system is compressed, scaling is observed for the elastic moduli, the coordination number, but also for the density of vibrational modes. In the present work we derive theoretically these results, and show that they apply to various systems such as granular matter and silica, but also to colloidal glasses. In particular we show that: (i) these systems present a large excess of vibrational modes at low frequency in comparison with normal solids, called the "boson peak" in the glass literature. The corresponding modes are very different from plane waves, and their frequency is related to the system coordination; (ii) rigidity is a non-local property of the packing geometry, characterized by a length scale which can be large. For elastic particles this length diverges near the jamming transition; (iii) for repulsive systems the shear modulus can be much smaller than the bulk modulus. We compute the corresponding scaling laws near the jamming threshold. Finally, we discuss the implications of these results for the glass transition, the transport, and the geometry of the random close packing.

1. Introduction

1.1 Anomalous properties of amorphous solids

In the last century, the development of statistical physics revolutionized our understanding of matter. It furnished a microscopic explanation of heat, and gave a description of different states of matter, such as the liquid or the solid state. Later, it explained that sudden transitions between these states can occur when a parameter is slowly tuned, despite the microscopic interactions staying the same. At the heart of these discoveries lie the concepts of equilibrium and entropy. At equilibrium, all the possible states with identical energy have an equal probability: this allows to define an entropy, and a temperature. Nevertheless, many systems around us are not at equilibrium. These can be open systems crossed by fluxes of matter and heat, such as biological systems. Another case is glassy systems, such as structural glasses or spin glasses, where the characteristic times become so slow that there are never equilibrated on experimental time scales. Finally there are also systems where particles are too large to be sensitive to temperature, such as granular matter. These systems are still poorly understood, and one of the current goal of nowadays statistical physics is to explain their original properties, and hopefully to find generic methods to describe them.

We do not have a satisfying description of amorphous systems such as structural glasses, colloids, emulsions or granular matter. This is particularly apparent when one considers the low temperature properties of glasses [1]. Their low-temperature specific heat has a nearly-linear temperature dependence rather than varying as T^3 as would be found in a crystal [1]. The prevailing explanation for this linear specific heat is in terms of tunneling in localized two-level systems [2]: atoms or group of atoms switch between two possible configurations by tunneling. This phenomenological model has also explain the T^2 dependence of the thermal conductivity at very low temperature. However, several empirical facts are still challenging the theory [3, 4]. Furthermore, after 30 years of research there is yet no accepted picture of what these two-levels systems are. At higher temperature, around

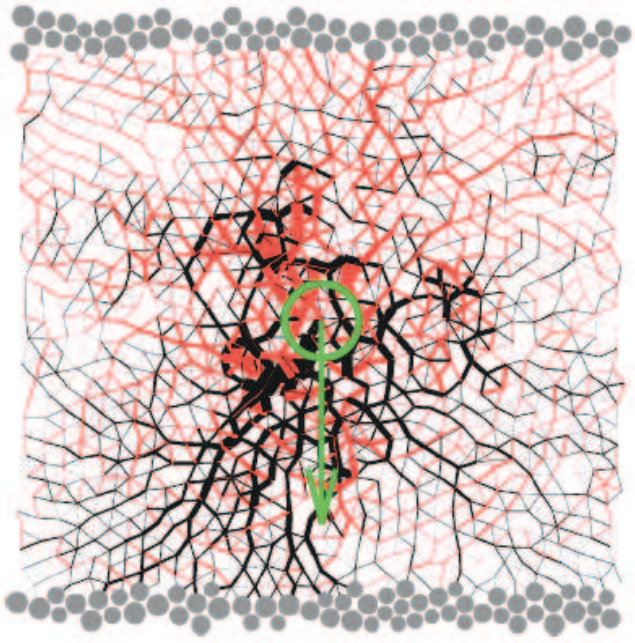

Fig. 1.1: Response to a monopole of force in a Lennard-Jones system of 1000 particles. Black (grey) lines correspond to compressive (tensile) stesses. Leonforte et al. [8]

typically 10 K which corresponds to the Thz frequency range for phonons, other universal properties of glasses are not fully understood. In particular the thermal conductivity displays a plateau, which suggests that at these frequencies phonons are strongly scattered. This effect is significant: for example in silica glass, the thermal conductivity is several orders of magnitude smaller than in the crystal of the same composition [5].

Athermal amorphous systems, such as granular matter, also display fascinating properties, both in their static behavior and in their rheology. The following puzzle underlines the subtlety of force propagation in granular matter [10]: the supporting force under a conical heap of poured sand is a minimum, rather than a maximum, at the center of the pile where it is deepest. As we shall discuss in the next Chapter, it has been proposed that in granular medium the force propagates differently than in a continuous elastic body [11, 12]. It turns out experimentally [13, 14] and numerically [15, 8] that an elastic-like behavior is recovered at large distances. Curiously enough, the cross-over length can be large in comparison with the particle size. Fig.(1.1) shows the response to a point force in a Lennard-Jones simulations [8] at zero temperature. The average response is similar to the one of a continu-

ous elastic medium, but near the source the fluctuations are of the order of the average. They decay exponentially with distance, with a characteristic length of roughly 30 particles sizes. One may ask what determines such a distance, below which an amorphous solid behaves as a continuous medium. More generally, what length scales characterize these systems?

The length scales we are discussing might also affect the rheology of granular matter. An interesting question is how grains flows, or how they compact [16]. For example if a layer of sand is inclined, an avalanche is triggered. Interestingly the angle θ of avalanche appears to be controlled by the width h of the granular layer. θ decreases when h grows when h is smaller than of the order of ten particle sizes. Similar length scales also appear in the spatial correlations of the velocities of grains in dense flows [17].

A particularity of the amorphous state is that it is not at equilibrium. Consequently the properties and the microscopic structure of these systems depend much on their history. For example if a granular pile is made by a uniform deposition, rather than by pouring sand from the top, the supporting force does not display the minimum discussed above at the center of the pile, but rather a flat maximum. Often amorphous solids are obtained from a fluid phase by varying some parameters such as temperature, density or applied shear stress until the system stops flowing: this is the jamming transition [18]. As the dynamics greatly slows down once this transition is passed, the structure of amorphous solids does not to differ too much from the marginally stable state at the transition. Thus a better understanding of the microscopic features of amorphous solids requires a better knowledge of the jamming mechanisms. It is a hard and much studied problem. When a glass is cooled rapidly enough to avoid crystallization, the relaxation times rapidly grow. In some cases the relaxation times follow an Arrhenius law with temperature; such glasses are called "strong". If the relaxation times grow faster, the glass is "fragile" [19]. There is no available theory to compute quantitatively the temperature dependence of the relaxation times, and to decide a priori which glasses are strong or fragile. Recently it was observed numerically and experimentally that the relaxation in the super-cooled is very heterogeneous and involves rearrangements of particles clusters [21, 20]. Althought several models of the glass transition predict such heterogeneities, see e.g. [22] and references therein, their cause and nature is still a much debated question.

Although they can lead to collective dynamics, most of the spatial models of the relaxation near the jamming threshold have purely *local* rules. This is the case for example for kinetically constrained models [23] where particles are allowed to move individually if their direct neighborhood satisfies some specific conditions. The starting point of the present work is the following remark: the stability against individual particle displacement is much less

demanding than the stability toward collective motions of particles. In d dimensions, $d+1$ neighbors are sufficient to pin one particle. As we shall discuss in details later, Maxwell showed that $2d$ contacts per particle *on average* are necessary to guarantee the stability of a solid [24]. The fact that the criterion of stability is non-local suggests that it is so for the minimal motions responsible for the relaxation. In any case, this underlines the importance of understanding what guarantees the stability of an assembly of interacting particles. In what follows we aim to furnish a microscopic description of the rigidity characterizing amorphous solids.

The informations about the rigidity of a solid against collective particle motions are contained in the density of the vibrational modes $D(\omega)$: a system is stable if there are no unstable modes. In a continuous isotropic elastic medium, the invariance by translation implies that the vibrational modes are plane waves. As a consequence in three-dimensions the density of vibrational modes $D(\omega)$ follows the Debye law $D(\omega) \propto \omega^2$. By contrast, at low frequency all glasses present an excess of vibrational modes in comparison with the Debye behavior. This excess of vibrational modes is the so-called "boson peak" [1] which appears as a maximum in $D(\omega)/\omega^2$. It is observed in particular in scattering experiments, see Fig.(1.2). The frequency of the peak lies in the terahertz range, that is typically between $\omega_D/10$ and $\omega_D/100$, where ω_D is the Debye frequency.

Several empirical facts suggest that the presence of these excess modes is related to many of the original properties of amorphous solids. In most glasses [27, 28, 29, 30], with some exceptions as silica [31], this boson peak shifts toward zero frequency when the glass is heated, as shown in Fig(1.2). Eventually the peak reaches zero frequency, as it has been observed numerically [32] and empirically [28]. This suggests that in some glasses the corresponding modes take part in the relaxation of the system [33, 34]. The presence of the boson peak also affects the low-temperature properties of glasses. The plateau in the thermal conductivity appears at temperatures that correspond to the boson peak frequency [1], which suggests that the excess-modes do not contribute well to transport. Furthermore, since these modes are soft, one expects that their non-linearities are important. Hence these modes may form two-levels systems [35, 36]. Finally, as the linear response to any force or deformation can be expressed in terms of the vibrational modes, it is reasonable to think that the boson peak affects force propagation. The recent simulations from which Fig.(1.1) is taken show that the length scale that appears in the response to a point force also appears in the normal mode

[1] The term "boson peak" was introduced because the amplitude of the scattering peak varies according to the Bose-Einstein factor at low temperature.

analysis: only for larger system sizes the lowest frequency modes are the one expected from a continuous elastic description [6, 7].

This excess of vibrational modes has been studied with diverse approaches. There are phenomenological models also dealing with two-levels systems such as the "soft potential theories" [35, 37, 38], which assumes the presence of strongly anharmonic localized soft potential with randomly distributed parameters. A second approach consists in studying the vibrations of elastic network with disorder. Simulations of a harmonic lattice with a random distribution of force constants [39, 40] exhibit a density of states qualitatively similar to what is observed with glasses in scattering experiments. From the theoretical point of view, models that assume spatially fluctuating elastic constants [41] show an excess of modes whose frequency decreases with the amplitude of the disorder. Recently further developments were proposed using the euclidean random matrix theory [42, 43, 44], where an assembly of particles at infinite temperature is considered. The density of states corresponds to the spectrum of a disordered matrix, the dynamical matrix. When the density ρ is infinite, the system behaves as a continuous medium. To approximate the density of states at finite ρ one uses perturbation theory in the inverse density. This leads to an excess of modes whose frequency goes to zero as the density decreases toward a finite threshold, and furnishes several exponents that describes the density of states at this transition. A third approach uses the mode coupling theory (MCT). MCT models the dynamics of supercooled liquids, and predicts a glass transition at finite temperature where the relaxation time diverges. In the glass region, the structure does not fully relax for any waiting time. Nevertheless if the mode coupling equations are used into this glass region, the dynamic that appears is non-trivial. It can be interpreted in terms of harmonic vibrations [45] around a frozen amorphous structure. The corresponding spectrum displays an excess of modes that converges toward zero frequency at the transition.

These different approaches describe the presence of an elastic instability, and make predictions on how the density of states behaves with some parameters when this instability is approached. Nevertheless they have several drawbacks. In these models, disorder is the main cause for the excess density of states. This is inconsistent with scattering data that show that some crystals also display excess vibrational modes [46, 47, 48, 49]. In particular silica, the glass with one of the strongest boson peak, as a density of states extremely similar to the crystals of identical composition and similar density, see Fig.(8.2), as we shall discuss in more details in Chapter 8. Thus the cause, and more importantly the *nature* of these excess modes are still unclear. In particular one may ask what microscopic features determine the vibrational properties of amorphous solids at these intermediate frequencies, and what

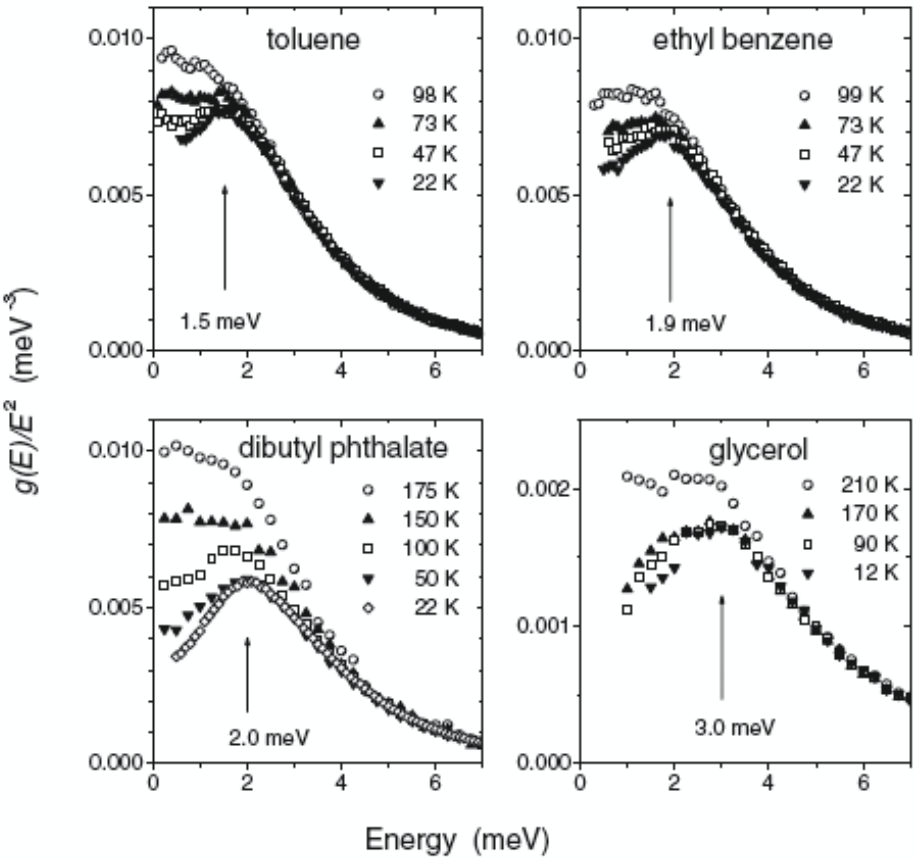

Fig. 1.2: Reduced density of states ($g(E) \approx D(\omega)$ following our notation) of collective motions in toluene, ethylbenzene, dibutylphthalate, and glycerol glasses. Arrows indicate the energy of the boson peak estimated from the data at lowest temperature. Chumakov et al., [30].

are the signatures of marginal stability at a microscopic level. In what follows we attempt to answer these questions in weakly-connected amorphous solids such as an assembly of repulsive, short-range particles. Then we argue that this description also applies to other systems, such as silica glass or particles with friction. Finally we derive some properties of colloidal glasses and discuss the possible implications of this approach for the glass transition.

1.2 Critical behavior at the jamming transition

Recently, C.S O'Hern, L.E Silbert et al. [50, 51] exhibited a system whose vibrational properties are dramatically different from a conventional solid. They simulate frictionless repulsive particles with short range interactions at zero temperature. The authors consider soft spheres. For inter-particle distance $r < \sigma$, the particles are in contact and interact with a potential:

$$V(r) = \frac{\epsilon}{\alpha}\left(1 - \frac{r}{\sigma}\right)^{\alpha} \tag{1.1}$$

where σ is the particle diameter and ϵ a characteristic energy. For $r > \sigma$ the potential vanishes and particles do not interact. Henceforth we express all distances in units of σ, all energies in units of ϵ, and all masses in units of the particle mass, m. The simulations were done for $\alpha = 2$ (harmonic), $\alpha = 5/2$ (Hertzian contacts) and $\alpha = 3/2$. When the packing fraction ϕ is low, such system is in a gas phase, and the pressure p is zero. At high packing fraction, it forms a solid and has a positive pressure. There is a transition between this two phases where the pressure vanishes: this is the jamming transition. At that point the density of states behaves as a *constant* instead of the quadratic dependence expected for normal solids, see Fig(1.3). The authors also study the solid phase when the pressure decreases toward zero. They find that the jamming transition acts as a *critical point*: the microscopic structure, the vibrational modes and the macroscopic elastic properties display scaling behaviors with the pressure p or with $\phi - \phi_c$, where ϕ_c is the packing fraction at the transition. In three dimensions $\phi_c \to 0.64$ which corresponds to the random close packing [2]. Concerning the structure, the coordination number z, which is the average number of contacts per particles, is found to follow:

$$z - z_c \sim (\phi - \phi_c)^{\frac{1}{2}} \tag{1.2}$$

[2] The parameter $\phi - \phi_c$ is somewhat less natural than the pressure because ϕ_c can vary from sample to sample. The distribution of ϕ_c converges to a well-defined value only when the number of particle N diverges. Nevertheless, the parameter $\phi - \phi_c$ has the advantage of being purely geometrical, and following [50] we should use it in most cases.

independently of the potential, where $z_c = 2d$, and d is the spatial dimension. This singular increase of the coordination was already noticed in [52]. Another striking observation is the presence of a singularity in the pair correlation function $g(r)$ at the jamming threshold. $g(r)$ has an expected delta function of weight z_c at a distance 1 that represents all particles in contact. But it also displays the following singularity:

$$g(r) \sim \frac{1}{\sqrt{r-1}} \tag{1.3}$$

which indicates that there are many particles *almost* touching. Again this is independent of the potential. This property was observed in other situations [53]. Note that Eq.(1.2) and (1.3) are related. A small affine compression of the configuration is equivalent to an increase of the particle diameter of an amount $\approx \frac{\phi-\phi_c}{3\phi_c}$. At the jamming threshold this would lead to an increase in the coordination number:

$$z - z_c \approx 4\pi \int_1^{1+\frac{\phi-\phi_c}{3\phi_c}} g(r) r^2 dr \sim \int_1^{1+\phi-\phi_c} \frac{1}{\sqrt{r-1}} dr \sim (\phi-\phi_c)^{\frac{1}{2}} \tag{1.4}$$

as observed.

As we mentioned, these simulations also reveal unexpected features in the density of states, $D(\omega)$: (a) As shown in Fig(1.3), when the system is most fragile, at $p \to 0$, $D(\omega)$ has a plateau extending down to zero frequency with no sign of the standard ω^2 density of states normally expected for a three-dimensional solid. (b) As shown in the inset to that figure, as p increases, the plateau erodes progressively at frequencies below a frequency ω^*, which scales in the harmonic case as:

$$\omega^* \sim \delta z \tag{1.5}$$

(c) The value of $D(\omega)$ in the plateau is unaffected by this compression. (d) At frequency much lower than ω^*, $D(\omega)$ still increases much faster with ω than the quadratic Debye dependence.

Finally, it was conjectured by Alexander [36] that the elastic moduli should scale at the jamming transition. Such scaling properties where observed in emulsions near the jamming transition [54], but the presence of noise in the measure of the packing fraction makes the exponent hard to identify. In [50], the bulk modulus B, the shear modulus G and the pressure are found to follow:

$$B \sim (\phi-\phi_c)^{\alpha-2} \tag{1.6}$$

$$G \sim (\phi-\phi_c)^{\alpha-3/2} \tag{1.7}$$

$$p \sim (\phi-\phi_c)^{\alpha-1} \tag{1.8}$$

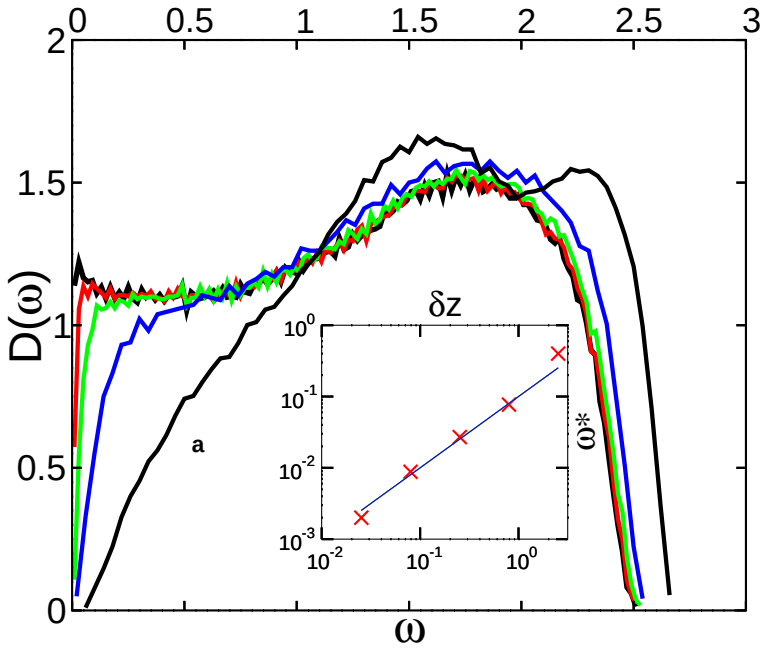

Fig. 1.3: $D(\omega)$ *vs* angular frequency ω for the simulation of Ref [50]. 1024 spheres interacting with repulsive harmonic potentials were compressed in a periodic cubic box to volume fraction ϕ, slightly above the jamming threshold ϕ_c. Then the energy for arbitrary small displacements was calculated and the dynamical matrix inferred. The curve labeled a is at a relative volume fraction $\phi - \phi_c = 0.1$. Proceeding to the left the curves have relative volume fractions 10^{-2}, 10^{-3}, 10^{-4}, 10^{-8}, respectively. Inset: Scaling of ω^* *vs* δz. ω^* for each $(\phi - \phi_c)$ is determined from the data in the main panel as the frequency where $D(\omega)$ is half of the plateau value. δz vs $(\phi - \phi_c)$ is obtained from the scaling measured in [50]. The line has slope 1.

These results raise many questions. Among others: (i) the vibrations of a normal solid are plane waves; what are the vibrations of a random close packing of elastic spheres? (ii) How is the behavior of the structure and the density of states related? This critical behavior is a stringent test to such theories. (iii) How does the microscopic structure, for example the coordination number, depend on the the system history? (iv) What are the different elastic properties of this system, that behaves almost as a liquid at the transition, as the shear modulus becomes negligible compare to the bulk modulus? For example, how does it react to a local perturbation?

1.3 Organization of the thesis

At the center of our argument lies the concept of *soft modes*, or *floppy modes*. These are collective modes that conserve the distance at first order between any particles in contact. They have been discussed in relation to various weakly-connected networks such as covalent glasses [55, 56], Alexander's models of soft solids [36], models of static forces in granular packs[57, 60] and rigidity percolation models, see e.g. [61]. As we will discuss below, they are present when a system is not enough connected. As a consequence, as Maxwell showed [24], a system with a low average coordination number z has some soft modes and therefore is not rigid. There is a threshold value z_c where a system can become stable, such a state is called *isostatic*. As we shall discuss later, this is the case at the jamming transition, if rattlers (particles with no contacts) are excluded. There are no zero-frequency modes except for the trivial translation modes of the system as a whole. However, if any contact were to be removed, there would appear one soft mode with zero frequency. Using this idea we will show in what follows that isostatic states have a constant density of states in any dimensions. When $z > z_c$, the system still behaves as an isostatic medium at short length scale, which leads to the persistence of a plateau in the density of states at high frequency.

The second concept we use is at the heart of the work of Alexander on soft solids [36]. In continuum elasticity the expansion of the energy for small displacements contains a term proportional to the applied stress (that we shall also call *initial stress* term following [36]), as we shall discuss in the next Chapter. It is responsible for the vibrations of strings and drumheads, but also for inelastic instability such as the buckling of thin rods. Alexander pointed out that this term has also strong effects at a microscopic level in weakly-connected solids. For example, it confers rigidity to gels, even though these do not satisfy the Maxwell criterion for rigidity. We will show that althought this term does not affect much the plane waves, it strongly affects

the soft modes. In a repulsive system of spherical particles it lowers their frequency. We shall argue that this can change dramatically the density of states at low frequency, as it will be confirmed by a comparison of simulations where the force in any contact is present, or set to zero. We show that these considerations lead to a inequality between the excess connectivity $\delta z \equiv z - z_c$ and the pressure that guarantees the rigidity of such amorphous solids. This relation between stability and structure will enable us to discuss how the history affects the microscopic structure of the system. In particlar, we shall argue that the preparation of the system used in [50] leads to a *marginally stable* state, even when $\phi > \phi_c$. This will account for both the scaling of the coordination, and for the divergence of the first peak in $g(r)$ at the random close packing.

A distinct and surprising property of the system approaching the jamming threshold is the nature of the quasi-plane waves that appear at lower frequency than the excess of modes. The peculiar nature of the transverse waves already appears at zero wave vector, as the shear modulus becomes negligible compared to the bulk modulus near jamming. If the response to a shear stress were be a perfect affine displacement of the system, the corresponding energy would be of the same order of the energy induced by a compression. As this is not the case, this indicates the presence of strong non-affine displacements in the transverse plane waves. To study this problem we shall introduce a formalism that writes the responses of the system in terms of the force fields that balance the force on every particle. This enables us to derive the scaling of the elastic moduli, and to compute the response to a local perturbation at the jamming threshold. We show that this response extends in the whole system.

We study how these ideas apply to real physical systems, such as granular matter, glasses and dense colloidal suspensions. In granular matter friction is always present. We derive the equation of the soft modes with friction. The main difference with frictionless particles is that rotational degrees of freedom of grains now matter, but our results on the vibrational modes and on the elastic properties are unchanged. Then we discuss the case of glasses. In these systems the coordination number is not well defined, as there are long range interactions such as Van der Waals forces. We show that if the hierarchy of interactions strengths is large enough, our description of the boson peak still applies. In particular we argue that the boson peak of silica glass corresponds to the slow modes that appear in weakly connected systems. Our argument also rationalizes why the crystal of identical composition and similar density, the crystobalite, has a similar density of state. We propose testable predictions to check if the same description holds for Lennard-Jones systems. Finally we study dense hard sphere liquids. Our main achievement

is to derive an effective potential that describes the hard sphere interaction when the fast temporal fluctuations are averaged out. Our effective potential is exact at the jamming threshold. It allows to define normal modes and to derive several properties of a hard sphere glass. In particular it implies that the jamming threshold act as a critical point both in the liquid and in the solid phases. This suggests original relaxation processes.

The thesis is organized as follows. Chapter 2 is introductive: we define rigidity and soft modes and discuss how these concept were used to study covalent glasses, force propagation or gels. In the Chapter 3 we use a simple geometric variational argument based on the soft modes to show that isostatic states have a constant density of states. The argument elucidates the nature of these excess-modes.In the Chapter 4 we compute the density of states when the coordination of the system $z = z_c+\delta z$ increases with the packing fraction. At that point we neglect the effect of the applied stress on the vibrations. This approximation corresponds to a real physical system: a network of relaxed springs. We show that such system behaves as an isostatic state for length scales smaller than $l^* \sim \delta z^{-1}$. This leads to a plateau in the density of states for frequency higher than $\omega^* \sim \delta z$. At lower frequency, we expect that the system behaves as a continuous medium with a Debye behavior, which is consistent with our simulations. We extend this result to the case of tetrahedral networks. In Chapter 5 we study the effect of the applied pressure on $D(\omega)$. We show that althought it does not affect much the plane waves, the applied stress lowers the frequency of the anomalous modes. We give a simple scaling argument to evaluate this effect, and we discuss its implication for the density of states. Incidentally this also furnishes an inequality between δz and the pressure which generalizes the Maxwell criterion for rigidity. We discuss the different length scales that appear in the problem. In Chapter 6, we discuss the influence of the cooling rate and the temperature history on the spatial structure and the density of states of the system. We show that the scaling of the coordination and the divergence in $g(r)$ are related to the marginal stability of the system of [50]. Some elastic properties of this tenuous system are computed in Chapter 7 , in particular the elastic moduli. Chapter 8 is devoted to the applications of our arguments to granular matter and glasses. This approach explains the qualitative shape of the density of state of silica. In Chapter 9 we derive an effective potential for hard spheres, and compute some properties of a hard spheres liquid near the glass transition. To conclude in chapter 11 we discuss the possible applications of these ideas to the low temperature properties of glasses, the glass transition and the rheology of granular matter.

2. Soft Modes and applications

2.1 Rigidity and soft modes

More than one century ago, Maxwell [24], working on the stability of engineering structures, studied the necessary conditions for the rigidity of an assembly of interacting objects. His response is as follows: consider for example a network of N point particles connected with N_c relaxed springs of stiffness unity in a space of dimension d. The expansion of the energy E is:

$$\delta E = \frac{1}{2}\sum_{\langle ij\rangle}[(\delta\vec{R_j} - \delta\vec{R_i}).\vec{n}_{ij}]^2 \tag{2.1}$$

where the sum is taken on every couple of particles in contact $\langle ij\rangle$, $\vec{n}_{ij}$ is the unit vector going from i to j, and $\delta\vec{R_i}$ is the displacement of particle i. It is convenient to express Eq.(2.1) in matrix form, by defining the set of displacements $\delta\vec{R}_1...\delta\vec{R}_N$ as a dN-component vector $|\delta\mathbf{R}\rangle$. Then Eq. (2.1) can be written in the form:

$$\delta E = \langle\delta\mathbf{R}|\mathcal{M}|\delta\mathbf{R}\rangle \tag{2.2}$$

The corresponding matrix $\mathcal{M}$ is known as the dynamical matrix [62], see footnote[3] for an explicit tensorial notation. The $3N$ eigenvectors of the dynamical matrix are the normal modes of the particle system, and its eigenvalues are the squared angular frequencies of these modes. A system is rigid if it has no *soft mode*, which are the modes with zero energy. Since Eq.(2.1) is a sum of positive terms, such modes satisfy:

$$(\delta\vec{R_i} - \delta\vec{R_j}).\vec{n}_{ij} = 0 \text{ for all } N_c \text{ contacts } \langle ij\rangle \tag{2.3}$$

This linear equation defines the vector space of displacement fields that conserve the distances at first order between particles in contact. The particles can yield without restoring force if their displacements lie in this vector space. Fig(2.1) furnishes an example of such mode. Note that Eq.(2.3) is purely geometrical and does not depend on the interaction potential.

[3] $\mathcal{M}$ can be written as an N by N matrix whose elements are themselves tensors of rank d, the spatial dimension $\mathcal{M}_{ij} = -\frac{1}{2}\delta_{\langle ij\rangle}\vec{n}_{ij}\otimes\vec{n}_{ij} + \frac{1}{2}\delta_{i,j}\sum_{<l>}\vec{n}_{il}\otimes\vec{n}_{il}$ where $\delta_{\langle ij\rangle} = 1$ when i and j are in contact, the sum is taken on all the contacts l with i.

Fig. 2.1: Illustration of a rigid and a floppy network made of 4 particles interacting with springs. $Nd-d(d+1)/2 = 5$ contacts are required to rigidify the structure. The red arrow of the floppy network indicates the soft mode that appears when one contact from the rigid system is removed.

Maxwell noticed that Eq.(2.3) has N_c constraints and $Nd - d(d+1)/2$ degrees of freedom if we substract global translations and rotations. Each equation restricts the $dN - d(d+1)/2$-dimensional space of $|\delta\mathbf{R}\rangle$ by one dimension. In general, these dimensions are independent, so that the number of independent soft modes is $dN - d(d+1)/2 - N_c$. A rigid system must not have any soft modes, and therefore as at least as many constraints as it has degrees of freedom. For a large system this yields for the average coordination $z \equiv 2N_c/N$:

$$z \geq 2d \tag{2.4}$$

This is the Maxwell criterion for rigidity. It is important to note that it is a *global* criterion, as it discusses the stability toward collective motions of particles. A local criterion that treat the motions of single particles only leads to $z > d+1$. As we shall see, some systems live on the bound of Eq.(2.4), they are called *isostatic*.

If this criterion has been known for quite a long time, it is only in the last decades that the concept of soft mode of non-rigid systems was used to study gels, covalent glasses and force propagation in granular matter. In the present Chapter we discuss these ideas.

2.2 Soft modes and force propagation

Recently several theories were proposed to describe the response to forces in sand. Some [11, 12] have argued that granular matter requires a new constitutive law: they postulate a linear relation between the components of

the stress tensor, the "null stress law". This leads to a continuum mechanics different from elasticity, where forces *propagate* along favored directions. In experiments, this description breaks down on large length scales [13, 14] where an elastic-like behavior is recovered. Nevertheless, as we shall see now, this theory was further justify for frictionless grains [57, 58, 59] using the concept of soft modes.

There is an obvious connection between soft modes and forces: soft modes do not have restoring force. Therefore a non-rigid system can resist to an external force $|\mathbf{F}\rangle \equiv \{\vec{F}_i\}$, where i labels the particles, only if the external force is orthogonal to each soft mode β, that is:

$$\langle \mathbf{F}|\delta\mathbf{R}^\beta\rangle \equiv \sum_i \vec{F}_i \cdot \delta\vec{R}_i^\beta = 0 \tag{2.5}$$

If Eq.(2.5) is not satisfied, the system yields along the soft modes, which are thus the directions of fragility of the system.

To apply this idea to granular matter, the starting point is the following remark: an assembly of frictionless hard spheres (or equivalently elastic spheres at the jamming transition where the pressure vanishes) is exactly isostatic, as was shown in particular in [57, 65, 60] and confirmed in the simulation of [50]. The argument for hard spheres is as follows: on the one hand, it is a rigid system and therefore must satisfy the bound of (2.4). On the other hand, the distance between hard spheres in contact must be equal to the diameter $\sigma = 1$ of the spheres:

$$||\vec{R}_i - \vec{R}_j|| = 1 \tag{2.6}$$

Eq.(2.6) brings exactly N_c constraints on the positions of the centers of the particles. Once again, there are $Nd - d(d+1)/2$ degrees of freedom for the particle positions. Therefore one must have $Nd - d(d+1)/2 \geq N_c$ which implies $z \leq 2d$ (note that this argument is contradicted in the case of the crystal whose coordination is larger. In the crystal, the constraints on the particle positions are redundant. Adding an infinitesimal poly-dispersity destroys this effect). Finally these two bounds lead to $z = 2d$.

The second remark is that an isostatic state is marginally rigid: if one contact is removed, one soft mode appears. This has the following consequence: an isostatic state is very sensitive to boundary conditions. Consider a subsystem of size L in a large isostatic system. Let $N_{ext} \sim L^{d-1}$ be the number of contacts of this subsystem with external beads. The number of contacts inside the subsystem N_{int} is on average $\langle N_{int}\rangle = Nd - N_{ext}/2$, where the factor $\frac{1}{2}$ shows up because a contact is shared by two particles. This implies that if the N_{ext} contacts were to be removed, the subsystem would not

be rigid. On average it will have $N_{ext}/2-d(d+1)/2$ soft modes. Consider now the force field composed of the N_{ext} contact forces applied by the external beads on the subsystem. It must be orthogonal to the $N_{ext}/2-d(d+1)/2$ soft modes. This implies that roughly half of the external contact forces are free degrees of freedom. If the contact forces are imposed on half of the boundary of the subsystem, the contact forces of the other half of the boundary are determined. This is very different from an elastic body where one can impose any stress on the whole boundary and compute the response of the system. In an isostatic system forces *propagate* from one side of the system to the other. In [57], the authors discuss the nature of the soft modes of isostatic anisotropic system. Using Eq.(2.5) they infer a null-stress law among the components of the stress tensor. This leads to the hyperbolic equations proposed to describe force propagation in granular matter [11, 12].

2.3 Covalent glasses

Phillips [55] used the soft modes, or "floppy modes", to study the structure of covalent glasses. The counting of degrees of freedom is slightly different from a network of springs where only the stretching of the contacts matters. Covalent interactions also display multi-body forces: the energy of the system depends on the angles formed by the different covalent bonds of an atom. These extra-terms in the energy, the "bond bending energy", bring extra-constraints on the soft modes. How many total constraints are there per atom of valence v? As we discussed with Eq.(2.3), the stretching of the v bonds leads to $v/2$ constrains on the soft modes (as there is one constraint per bond and each bond is shared by two atoms). Furthermore, the covalent bonds form $v(v-1)/2$ angles, each of them corresponds to a term in the energy expansion. Therefore the total number of constraints is $v/2 + v(v-1)/2 = v^2/2$.

Consider, following [55], chalcogenide alloys such as Ge_xSe_{1-x}, where the relative concentration x of Ge can vary from 0 to 1. Ge has a valence of 4 whereas Se has a valence of 2. When $x = 0$ the glass as a polymeric structure. When x increases the connectivity of the covalent network increases too, until it becomes rigid. This takes place when $x \cdot 4^2 + (1-x) \cdot 2^2 = 2d = 6$, that is when $x_c \approx 0.16$.

However, although the covalent network is floppy for smaller x, the glass is still rigid at lower concentration. In particular the shear modulus does not vanish below x_c [63]. The reason is that this counting argument neglects the weaker interactions induced by the lone pair electrons, that lead to short range repulsion and long range attraction (Van der Waals interaction). Nev-

ertheless several experimental results show that the transition at $x = 0.16$ affects certain properties of the system. In Ge-As-Se glasses the fragility, that quantifies how the dependence of the viscosity with temperature is different from an Arrhenius law, is maximal at the transition [64]. Furthermore, it was argued [55] that the composition of the best glass former should be x_c, as it appears experimentally, see Fig(2.2). A qualitative argument is as follows: when x increases toward x_c, the covalent network becomes more and more intricate and the viscosity of the system increases. It takes therefore more and more time to nucleate the crystal. On the other hand, if x increases above x_c, the covalent network is over-constrained, the bonds are frustrated and have to store energy. The configuration of the crystal, where the bonds can organize to avoid this frustration, becomes more and more favorable with x in comparison with the amorphous state, and is thus easier to nucleate.

This description is "mean field" as it does not consider the possible spatial fluctuations of the coordination. Such fluctuations may lead to a system where rigid, high-coordinated regions coexist with floppy, weakly-coordinated regions. To study this possibility models were proposed such as rigidity percolation, see e.g. [61] and reference therein. In its simplest form, this model considers springs randomly deposited on a lattice. When the concentration of springs increases, there is a transition when a rigid cluster percolates in the system almost fully floppy. Such a cluster is a non-trivial fractal object and contains over-constrained regions. In this form, this model is at infinite temperature, as there are no correlation among the contacts deposited. It is interesting to note the difference with the transition of jamming that we study in what follows, which is also a transition of rigidity. At the jamming threshold the correlations among particles are obviously important, as the temperature is not infinite. The system is exactly isostatic, as we shall discuss, and does not contain over-constrained regions. It is a normal d-dimensional object with only very few holes of size of order unity [4], the "rattlers", which are isolated particles without contact [50].

[4] The absence of large voids, or rattlers, has geometrical origins. Because particles are repulsive, the rigid region surrounding a void must be convex. A large void necessitates the creation of a vault. A flat vault is impossible as forces cannot be balanced on any particle of the surface. A weakly curved vault imposes drastic constraints on the distribution of angles of contact between these particles. This suggests that the probability of making a large void decays very fast, presumably at least exponentially in the surface of the void. If friction is present such voids might form more easily.

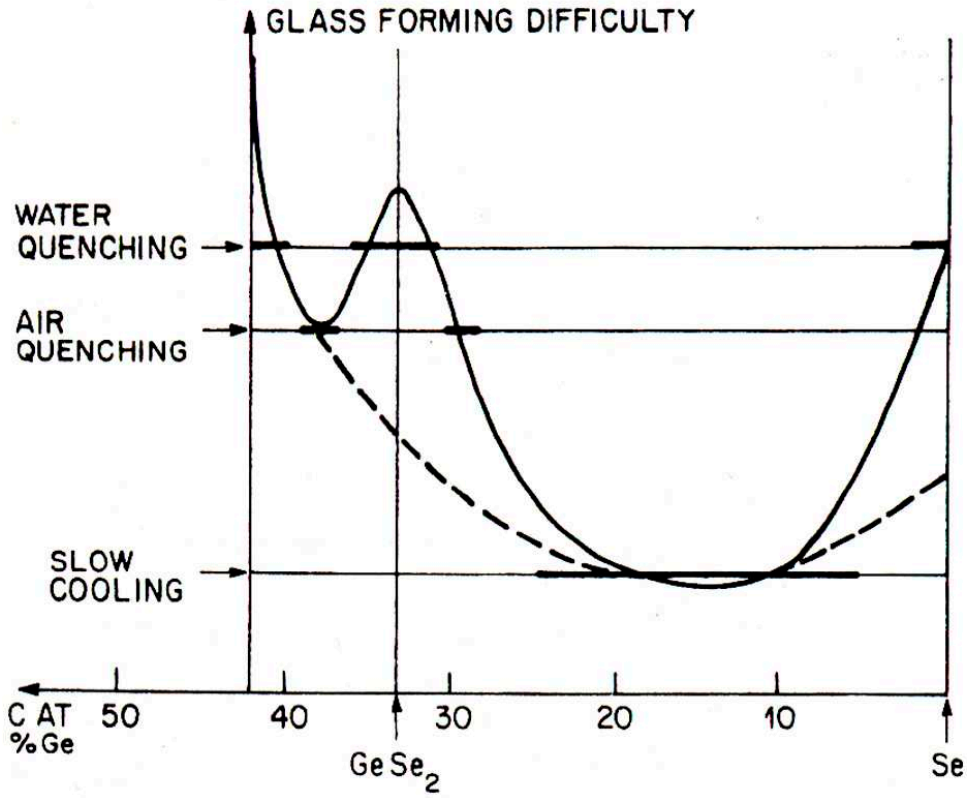

Fig. 2.2: The quenching rate (or difficulty of glass formation) was plotted as a function of x in Ge_xSe_{1-x} alloys. Both lines are sketched in to guide the reader's eye. The dashed line corresponds to the network effect discussed in the text. Phillips [55].

2.4 The rigidity of "soft solids"

Alexander noticed [36] the following contradiction: there are solids less connected than required by the Maxwell criterion, and that are nevertheless rigid. For example, one may describe a gel as an assembly of reticulated point linked by springs that are the polymers. In general the coordination of such system is less than 6. Why then a gel has a finite shear modulus? The starting point of Alexander's answer is the simple following remark: the stress influences the frequency of the vibrations. At a macroscopic level the presence of a negative stress (that is, when the system is stretched) is responsible for the fast vibrations of strings or drumheads. When a system is compressed, the stress is positive and lowers the frequency of the modes that can even become unstable, as for the buckling of a thin rod. To discuss the role of stress at a microscopic level, consider particles interacting with a potential $V(r)$. The expansion of the energy leads to:

$$\delta E = \sum_{ij} V'(r_{ij}^{eq}) dr_{ij} + \frac{1}{2} V''(r_{ij}^{eq}) dr_{ij}^2 + O(dr_{ij}^3) \tag{2.7}$$

where the sum is over all pairs of particles, r_{ij}^{eq} is the equilibrium distance between particles i and j. In order to get an expansion in the displacement field $\delta\vec{R}_i$ we use:

$$dr_{ij} = (\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij} + \frac{[(\delta\vec{R}_j - \delta\vec{R}_i)^{\perp}]^2}{2r_{ij}^{eq}} + O(\delta\vec{R}^3) \tag{2.8}$$

where $(\delta\vec{R}_j - \delta\vec{R}_i)^{\perp}$ indicates the projection of $\delta\vec{R}_j - \delta\vec{R}_i$ on the plane orthogonal to $\vec{n}_{ij}$. When used in Eq.(2.7), the linear term in the displacement field disappears (the system is at equilibrium) and we obtain:

$$\delta E = \{\sum_{ij} V'(r_{ij}^{eq}) \frac{[(\delta\vec{R}_j - \delta\vec{R}_i)^{\perp}]^2}{2r_{ij}^{eq}}\} + \frac{1}{2} V''(r_{ij}^{eq}) [(\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij}]^2 + O(\delta\vec{R}^3) \tag{2.9}$$

The difference from Eq.(2.1) is the term inside curly brackets. Such term is called the "initial stress" term in [36] as it is directly proportional to the forces $V'(r_{ij}^{eq})$. We shall also refer to it as the applied stress term. If the system has a negative pressure p this term increases the frequency of the modes. As a consequence if $p < 0$ all the soft modes of a weakly-connected system gain a finite positive energy: the system is rigid. This occurs in gels: the osmotic pressure of the solvent is larger than the external pressure. This imposes that the network of reticulated polymers carry a negative pressure to compensate this difference: the polymers are stretched. This rigidifies the

system. As a consequence, the shear modulus of gels is directly related to the osmotic pressure of the solvent.

3. Vibrations of isostatic systems

3.1 Isostaticity

When a system of repulsive spheres jams at zero temperature, the system is isostatic [57, 60, 65] (when the rattlers, or particles without contacts, are removed). As we said above, on the one hand it must be rigid and satisfy the bound of (2.4). One the other hand, it cannot be more connected: this would imply that the contacts are frustrated as they cannot satisfy Eq.(2.6). That is, there would not be enough displacements degrees of freedom to allow particles in contact to touch each other without interpenetrating, as we discussed for hard spheres in the last Chapter. Thus the energy of the system, and the pressure, would not vanish at the transition. It must be the case, since an infinitesimal pressure can jam an athermal gas of elastic particles.

In terms of energy expansion, since the pressure vanishes at the transition, the initial stress in bracket in Eq.(2.9) vanishes. For concreteness in the following Chapters we consider the harmonic potential, corresponding to $\alpha = 2$ in Eq.(1.1). The expansion of the energy is then given by Eq.(2.1). In Chapter 7 we generalize our findings to other soft sphere potentials and other types of interactions.

An isostatic system is marginally stable: if q contacts are cut, a space of soft modes of dimension q appears. For our argument below we need to discuss the extended character of these modes. In general when only one contact $\langle ij \rangle$ is cut in an isostatic system, the corresponding soft mode is not localized near $\langle ij \rangle$. This comes from the non-locality of the isostatic condition that gives rise to the soft modes; and was confirmed in the isostatic simulations of Ref [57], which observed that the amplitudes of the soft modes spread out over a nonzero fraction of the particles. This shall be proved by the calculation of Chapter 7 that shows that in an isostatic system, the response to a local strain does not decay with the distance from the source. When many contacts are severed, the extended character of the soft modes that appear depends on the geometry of the region being cut. If this region is compact many of the soft modes are localized. For example cutting all the contacts inside a sphere totally disconnects each inner particle. Most of the

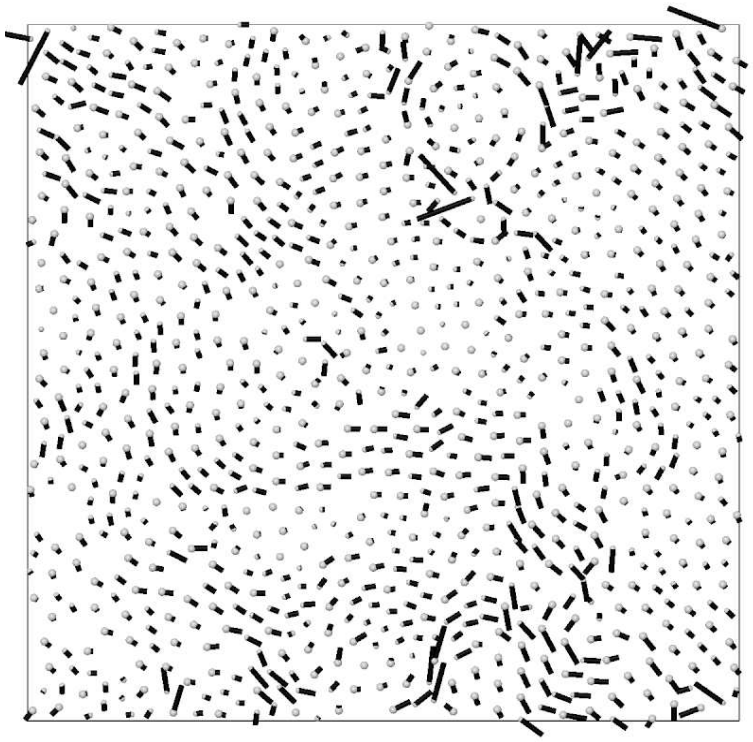

Fig. 3.1: One soft mode in two dimensions for $N \approx 1000$ particles. The relative displacement of the soft mode is represented by a line segment extending from the dot. The mode was created from a previously prepared isostatic configuration, periodic in both directions, following [50]. 20 contacts along the vertical edges were then removed and the soften modes determined. The mode pictured here is an arbitrary linear combination of these modes.

soft modes are then the individual translations of these particles and are not extended throughout the system.

In what follows we will be particularly interested in the case where the region of the cut is a hyper-plane as illustrated in Fig.(3.2). In this situation occasionally particles in the vicinity of the hyper-plane can be left with less than d contacts, so that trivial localized soft modes can also appear. However this represents only a finite fraction of the soft modes. We expect that there is a non-vanishing fraction q' of the total soft modes that are not localized near the hyper-plane. Rather, as when a single contact is cut, these modes should extend over the whole system, like the mode shown in Fig.(3.1). We shall define extended modes more precisely in the next section.

3.2 Variational procedure

We aim to show first that the density of states of an isostatic system does not vanish at zero frequency. Since $D(\omega)$ is the total number of modes per unit volume per unit frequency range, we have to show that there are at least of the order of ωL^d normal modes with frequencies smaller than ω for any small ω. As we justify later, if proven in a system of size L for $\omega \sim \omega_L \sim 1/L$, this property can be extended to a larger range of ω independent of L. Therefore it is sufficient to show that there are of the order of L^{d-1} normal modes with frequency of the order of $1/L$, instead of the order of one such mode in a continuous solid: the whole translation of the system. To do so we use a variational argument: $\mathcal{M}$ is a positive symmetric matrix. Therefore if a normalized mode has an energy δE, we know that the lowest eigenmode has a frequency $\omega_0 \equiv \sqrt{E_0} \leq \sqrt{\delta E}$. Such argument can be extended to a set of modes [5]: if there are m *orthonormal* trial modes with energy $\delta E \leq \omega_t^2$, then there are at least $m/2$ *normal* modes with frequency smaller than $\sqrt{2}\omega_t$. Therefore we are led to find of the order of L^{d-1} trial orthonormal modes with energy of order $1/L^2$.

3.3 Trial modes

For concreteness we consider the three-dimensional cubic N-particle system $\mathcal{S}$ of Ref [50] with periodic boundary conditions at the jamming threshold. We label the axes of the cube by x, y, z. $\mathcal{S}$ is isostatic, so that the removal of n contacts allows exactly n displacement modes with no restoring force. Consider for example the system $\mathcal{S}'$ built from $\mathcal{S}$ by removing the $q \sim L^2$ contacts crossing an arbitrary plane orthogonal to (ox); by convention at $x = 0$, see Fig.(3.2). $\mathcal{S}'$, which has a free boundary condition instead of periodic ones along (ox), contains a space of soft modes of dimension q [6], instead of one such mode —the translation of the whole system— in a normal solid. As stated above, we suppose that a subspace of dimension $q' \sim L^2$ of these soft modes contains only extended modes. We define the *extension* of a mode relative to the cut hyper-plane in terms of the amplitudes of the mode at distance x from this hyper-plane. Specifically the extension

[5] If m_α is the α'th lowest eigenvalue of $\mathcal{M}$ and if e_α is an orthonormal basis such that $\langle e_\alpha|\mathcal{M}|e_\alpha\rangle \equiv n_\alpha$ then the variational bound of A. Horn [Am. J. Math **76** 620 (1954)] shows that $\sum_1^q m_\alpha \leq \sum_1^q n_\alpha$. Since $qn_q \geq \sum_1^q n_\alpha$, and since $\sum_1^q m_\alpha \geq \sum_{q/2}^q m_\alpha \geq (q/2)m_{q/2}$, we have $qn_q \geq (q/2)m_{q/2}$ as claimed.

[6] The balance of force can be satisfied in $\mathcal{S}'$ by imposing external forces on the free boundary. This adds a linear term in the energy expansion that does not affect the normal modes.

e of a normalized mode $|\delta\mathbf{R}\rangle$ is defined by $\sum_i \sin^2(\frac{x_i\pi}{L})\langle i|\delta\mathbf{R}\rangle^2 = e$, where the notation $\langle i|\delta\mathbf{R}\rangle$ indicates the displacement of the particle i of the mode considered. For example, a uniform mode with $\langle i|\delta\mathbf{R}\rangle$ constant for all sites has $e = \frac{1}{2}$ independent of L. On the other hand, if $\langle i|\delta\mathbf{R}\rangle = 0$ except for a site i adjacent to the cut hyper-plane, the $x_i/L \sim L^{-1}$ and $e \sim L^{-2}$. We define the subspace of extended modes by setting a fixed threshold of extension e_0 of order 1 and thus including only soft modes β for which $e_\beta > e_0$. As we argued at the beginning of the Chapter, we expect that a fraction of the total soft modes are extended. Thus if q' is the dimension of the extended modes vector space, we shall suppose that q'/q remains finite as $L \to \infty$; i.e. a fixed fraction of the soft modes remain extended as the system becomes large. The appendix at the end of this Chapter presents our numerical evidence for this behavior.

We use them to build q' orthonormal trial modes of frequency of the order $1/L$ in the initial system $\mathcal{S}$. Let us denote $|\delta\mathbf{R}_\beta\rangle$ a normalized basis of the vector space of such extended modes, $1 \leq \beta \leq q'$. These modes are not soft in the jammed system $\mathcal{S}$ because they deform the previous q contacts located at $x = 0$, and therefore cost energy. Nevertheless a set of trial modes, $|\delta\mathbf{R}^*_\beta\rangle$, can still be formed by altering the soft modes so that they do not have an appreciable amplitude at the boundary where the contacts were severed. We seek to alter the soft mode to minimize the distortion at the severed contacts while minimizing the distortion elsewhere. Accordingly, for each soft mode β we define the corresponding trial-mode displacement $\langle i|\delta\mathbf{R}^*\rangle$ to be:

$$\langle i|\delta\mathbf{R}^*_\beta\rangle = C_\beta \sin(\frac{x_i\pi}{L})\langle i|\delta\mathbf{R}_\beta\rangle \tag{3.1}$$

where the normalization constant C_β depends on the spatial distribution of the mode β. If for example $\langle i|\delta\mathbf{R}\rangle = 0$ except for a site i adjacent to the cut plane, C_β grows without bound as $L \to \infty$. In the case of extended modes $C_\beta^{-2} \equiv \sum_{\langle ij\rangle} \sin^2(\frac{x_i\pi}{L})\langle j|\delta\mathbf{R}_\beta\rangle^2 = e_\beta > e_0$, and therefore C_β is bounded above by $e_0^{-\frac{1}{2}}$. The sine factor suppresses the problematic gaps and overlaps at the q contacts near $x = 0$ and $x = L$. Formally, the modulation by a sine is a linear mapping. This mapping is invertible if it is restricted to the extended soft modes. Consequently the basis $|\delta\mathbf{R}_\beta\rangle$ can always be chosen such that the $|\delta\mathbf{R}^*_\beta\rangle$ are orthogonal. Furthermore one readily verifies that the $|\delta\mathbf{R}^*_\beta\rangle$'s energies are small, because the sine modulation generates an energy of order $1/L^2$ as expected. Indeed we have from Eq.(2.1):

$$\delta E = C_\beta^2 \sum_{\langle ij\rangle} [(\sin(\frac{x_i\pi}{L})\langle i|\delta\mathbf{R}_\beta\rangle - \sin(\frac{x_j\pi}{L})\langle j|\delta\mathbf{R}_\beta\rangle) \cdot \vec{n}_{ij}]^2 \tag{3.2}$$

Using Eq.(2.3), and expanding the sine, one obtains:

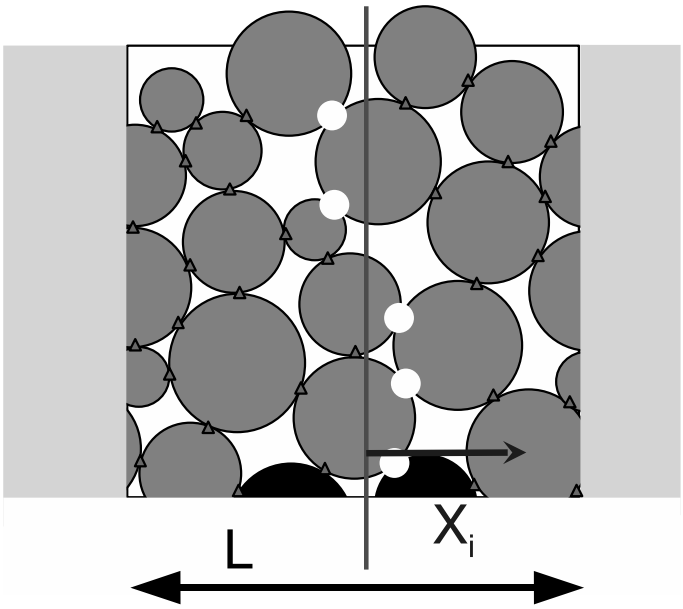

Fig. 3.2: Illustration of the boundary contact removal process described in the text. Eighteen particles are confined in a square box of side L periodically continued horizontally and vertically. An isostatic packing requires 33 contacts in this two-dimensional system. An arbitrarily drawn vertical line divides the system. A contact is removed wherever the line separates the contact from the center of a particle. Twenty-eight small triangles mark the intact contacts; removed contacts are shown by the five white circles.

$$\delta E \approx C_\beta^2 \sum_{\langle ij \rangle} \cos^2(\frac{x_i \pi}{L}) \frac{\pi^2}{L^2} (\vec{n}_{ij} \cdot \vec{e}_x)^2 (\langle j|\delta \mathbf{R}_\beta \rangle \cdot \vec{n}_{ij})^2 \tag{3.3}$$

$$\leq e_0^{-1} (\pi/L)^2 \sum_{\langle ij \rangle} \langle j|\delta \mathbf{R}_\beta \rangle^2 \tag{3.4}$$

where $\vec{e}_x$ is the unit vector along (ox). The sum on the contacts can be written as a sum on all the particles since only one index is present in each term. Using the normalization of the mode β and the fact that the coordination number of a sphere is bounded by a constant z_{max} ($z_{max} = 12$ for 3 dimensional spheres [7]), one obtains:

$$\delta E \leq e_0^{-1} (\pi/L)^2 z_{max} \equiv \omega_L^2 \tag{3.5}$$

One may ask if the present variational argument can be improved, for example by considering geometries of broken contacts different from the surface we considered up to now. When contacts are cut to create a vector space of extended soft modes, the soft modes must be modulated with a function that vanishes where the contacts are broken in order to obtain trial modes of low energy. One the one hand, cutting many contacts increases the number of trial modes. On the other hand, if too many contacts are broken, the modulating function must have many "nodes" where it vanishes. Consequently this function displays larger gradients and the energies of the trial modes increase. Cutting a surface (or many surfaces, as we shall discuss below) appears to be the best compromise between these two opposing effects. Thus our argument gives a natural limit to the number of low-frequency states to be expected.

Finally we have found of the order of L^2 trial orthonormal modes of frequency bounded by $\omega_L \sim 1/L$, and we can apply the variational argument mentioned above: the density of statesis bounded below by a constant below frequencies of the order ω_L. In what follows, the trial modes introduced in Eq.(3.1), which are the soft modes modulated by a sine, shall be called "anomalous modes".

[7] In a polydisperse system z_{max} could a priori be larger. Nevertheless Eq.(3.3) is a sum on every contact where the displacement of only one of the two particles appears in each term of the sum. The corresponding particle can be chosen arbitrarily. Chose the smallest particle of each contact. Thus when this sum on every contact is written as a sum on every particles to obtain Eq.(3.5), the constant z_{max} still corresponds to the monodisperse case, as a particle cannot have more contacts with particles larger than itself.

3.4 Argument extension to a wider frequency range

We may extend this argument to show that the bound on the initial density of states extends to a plateau encompassing a nonzero fraction of the modes in the system. If the cubic simulation box were now divided into m^3 sub-cubes of size L/m, each sub-cube must have a density of states equal to the same $D(\omega)$ as was derived above, but extending to frequencies of order $m\omega_L$. These subsystem modes must be present in the full system as well, therefore the bound on $D(\omega)$ extends to $[0, m\omega_L]$. We thus prove that the same bound on the average density holds down to sizes of the order of a few particles, corresponding to frequencies independent of L. Finally $D(\omega)$ does not vanish when $\omega \to 0$, as indicates the presence of the observed plateau in the density of states. We note that in d dimensions this argument may be repeated to yield a total number of modes, L^{d-1}, below a frequency $\omega_L \approx 1/L$, thus yielding a limiting nonzero density of states in any dimension.

We note that the trial modes of energy $\delta E \sim l^{-1}$ that we introduced by cutting in subsystems of size l are, by construction, localized to distance scale l. Nevertheless we expect these trial modes to hybridize with the trial modes of other subsystems, and the corresponding normal modes not to be localized on such length scale.

3.5 Appendix: Spatial distribution of the soft modes

In our argument we have assumed that when $q \sim L^{d-1}$ contacts are cut along a hyper-plane in an isostatic system, there is a vector space of dimension $q' = aq$ which contains only extended modes, such that a does not vanish when $L \to \infty$. A normalized mode $|\delta\mathbf{R}\rangle$ was said to be extended if $\sum_i \sin^2(\frac{x_i\pi}{L})\langle i|\delta\mathbf{R}\rangle^2 > e_0$, where e_0 is a constant, and does not depend on L. Here we show how to choose $e_0 > 0$ so that there is a non-vanishing fraction of extended soft modes. We build the vector space of extended soft modes and furnish a bound on its dimension.

Let us consider the linear mapping $\mathcal{G}$ which assigns to a displacement field $|\delta\mathbf{R}\rangle$ the displacement field $\langle i|\mathcal{G}\delta\mathbf{R}\rangle = \sin^2(x_i\pi/L)\delta\vec{R}_i$. For any soft mode $|\delta\mathbf{R}_\beta\rangle$ one can consider the positive number $a_\beta \equiv \langle\delta\mathbf{R}_\beta|\mathcal{G}|\delta\mathbf{R}_\beta\rangle \equiv \sum_i \sin^2(x_i\pi/L)\delta\vec{R}_{\beta_i}^{\,2}$. We build the vector space of extended modes by recursion: at each step we compute the a_β for the normalized soft modes, and we eliminate the soft mode with the minimum a_β. We then repeat this procedure

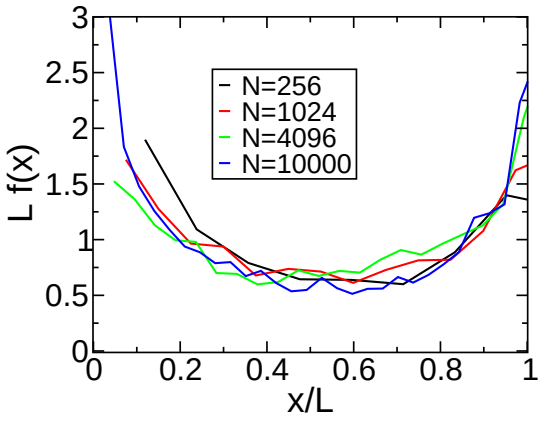

Fig. 3.3: The overlap function $f(x)$ as defined in Eq.(3.6) for different system sizes in three dimensions. Soft modes were created from isostatic configurations as described below Fig. 3.1.

in the vector space orthogonal to the soft modes eliminated. We stop the procedure when $a_\beta > e_0$ for all the soft modes β left. Then all the modes left are extended according to our definition. We just have to show that one can choose $e_0 > 0$ such that when this procedure stops, there are q' modes left, with $q' > aq$ and $a > 0$. In order to show that, we introduce the following overlap function:

$$f(x)dx \equiv q^{-1} \sum_{\beta=1,\dots,q} \sum_{x_i \in [x,x+dx]} [\delta \vec{R}_{i,\beta}]^2 \tag{3.6}$$

The sum is taken on an orthonormal basis of soft modes β and on all the particles whose position has a coordinate $x_i \in [x, x+dx]$. $f(x)$ is the trace of a projection, and is therefore independent of the orthonormal basis considered. $f(x)$ describes the spatial distribution of the amplitude of the soft modes. The $\delta\mathbf{R}_\beta$ are normalized and therefore:

$$\int_0^L f(x)dx = 1 \tag{3.7}$$

We have examined soft modes made from configurations at the jamming transition found numerically in [50]. The overlap function $f(x)$ was then computed for different system sizes L. These are shown in Fig. 3.3. It appears from Fig.(3.3) that i) when $f(x)$ is rescaled with the system size it collapses to a unique curve, and ii) this curve is bounded from below by a constant c ($c \approx 0.6$). Consequently one can bound the trace of $\mathcal{G}$: $tr\mathcal{G} = \int_0^L q f(x) \sin^2(x) > qc/2$. On the other hand one has $tr\mathcal{G} = \sum_{\beta=1}^{q} a_\beta$, where the sum is made on the orthonormal basis we just built. Introducing

the numbers a and e_0 such that there are $q' = aq$ extended modes, and using that the $a_\beta < 1$, one can bound this sum and obtain $tr\mathcal{G} < aq + (1-a)qe_0$. Choosing for example $e_0 = c/2 > 0$, one finds $a > c/(2-c)$, which is a constant independent of L as claimed.

4. Evolution of the modes with the coordination

4.1 An "isostatic" length scale

When the system is compressed and moves away from the jamming transition, the simulations showed that the extra-coordination number $\delta z \equiv z - z_c$ increases. In the simulation, the compression also creates forces on all contacts. In this Chapter we ignore these forces, and instead only consider the contact network created by compression, but in the absence of applied pressure. Any tension or compression in the contacts is removed. The effect on the energy is to remove the first bracketed term from Eq.(2.9) above, and the expansion of the energy is still given by Eq.(2.1). We note that removing these forces, which add to zero on each particle, does not disturb the equilibrium of the particles or create displacements. In this section we ignore the question of how δz depends on the degree of compression. We return to this question in the next section. Compression causes $\Delta N_c = N\delta z/2 \sim L^3\delta z$ extra constraints to appear in Eq.(2.1). Cutting the boundaries of the system, as we did above, relaxes $q \sim L^2$ constraints. For a large system, $q < \Delta N_c$ and Eq.(2.1) is still over-constrained so that no soft modes appear in the system. However, as the systems become smaller, the excess ΔN_c diminishes, and for L smaller than some $l^* \sim \delta z^{-1}$ where $\Delta N_c = q$, the system is again under-constrained, as was already noticed in [57]. This allows us to build low-frequency modes in subsystems smaller than l^*. These modes appear above a cut-off frequency $\omega^* \sim l^{*-1}$; they are the excess-modes that contribute to the plateau in $D(\omega)$ above ω^*. In other words, anomalous modes with characteristic length smaller than l^* are little affected by the extra contacts, and the density of states is unperturbed above a frequency $\omega^* \sim \delta z$. This scaling is checked numerically in Fig.4.1. It is in very good agreement with our prediction up to $\delta z \approx 2$.

At frequency lower than ω^* we expect the system to behave as a disordered, but not ill-connected, elastic medium, so that the vibrational modes are similar to the plane waves of a continuous elastic body. We refer to these

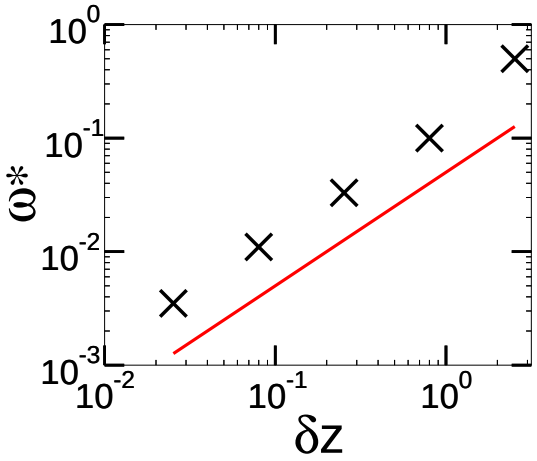

Fig. 4.1: ω^* as defined in the insert of Fig.1 *vs.* coordination, in the system with relaxed springs. The line as a slope one.

modes as "acoustic modes". Thus we expect $D(\omega)$ at small ω to vary as $\omega^2 c^{-3}$, where $c(\delta z)$ is the sound speed at the given compression. This c may be inferred from the bulk and shear moduli measured in the simulations; that we shall derive in chapter 7. One finds for the transverse velocity $c_t \sim (\delta z)^{\frac{1}{2}}$, and for the longitudinal velocity $c_l \sim \delta z^0$. Thus at low frequency the $D(\omega)$ is dominated by the transverse plane waves and at $\omega = \omega^*$ the acoustic density of states is $\omega^2 c_t^3 \sim \delta z^2 \delta z^{-3/2} \sim \delta z^{\frac{1}{2}}$: the acoustic density of states should be dramatically smaller than the plateau density of states. There is no smooth connection between the two regimes, thus we expect a sharp drop-off in $D(\omega)$ for $\omega < \omega^*$. Such drop-off is indeed observed, as seen in Fig(4.2). In fact, because of the finite size of the simulation, no acoustic modes are apparent at $\omega < \omega^*$ near the transition.

Thus the behavior of such system near the jamming threshold depends on the frequency ω at which it is considered. For $\omega > \omega^*$ the system behaves as an isostatic state: the density of states is dominated by anomalous modes. For $\omega < \omega^*$ we expect it to behave as a continuous elastic medium with acoustic modes. Since the transverse and the longitudinal velocities do not scale in the same way, the wavelengths of the longitudinal and transverse plane waves at ω^* are two distinct length scales l_l and l_t which follow $l_l \sim c_l \omega^{*-1}$ and $l_t \sim c_t \omega^{*-1}$. At shorter wave lengths we expect the acoustic modes to be strongly perturbed. Note that since $c_l \sim \delta z^0$, one has $l_l \sim l^*$. Interestingly, $l_t \sim \delta z^{-\frac{1}{2}}$ is the smallest system size at which plane waves can be observed: for smaller systems, the lowest frequency mode is not a plane wave, but an anomalous mode. l_t was observed numerically in [66].

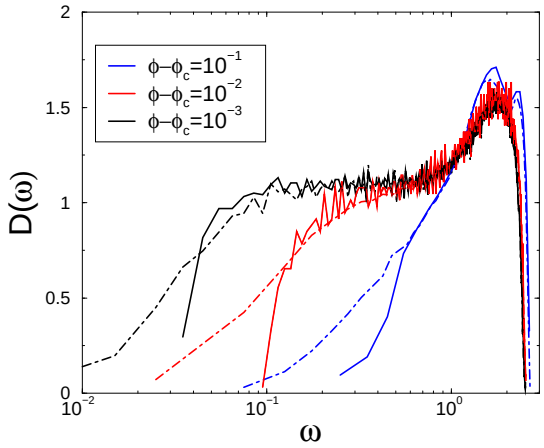

Fig. 4.2: Log-linear plot of the density of states for N=1024 for three values of $\phi - \phi_c$ in the soft spheres system (dotted line) and the system where the applied stress term has been removed (solid line).

4.2 Role of spatial fluctuations of z

Our argument ignores the spatial fluctuations of δz. If these fluctuations were spatially uncorrelated they would be Gaussian upon coarse-graining: then the extra number of contacts ΔN_c in a subregion of size L would have fluctuations of order $L^{d/2}$. The scaling of the contact number that appears in our description is $\Delta N_c \sim L^{d-1}$ and is therefore larger than these Gaussian fluctuations for $d > 2$. In other terms at the length scale l^* where soft modes appear, the fluctuations of the number of contacts inside the bulk are negligible in comparison with the number of contacts at the surface. Therefore the extended soft modes that are described here are not sensitive to fluctuations of coordination in three dimensions near the transition. In [67] we argued that in two dimensions there are spatial anti-correlations in z, and that the fluctuations do not affect the extended soft modes in two dimensions either.

Note that these arguments do not preclude the existence of low-frequency localized modes that may appear in regions of small size $l \ll l^*$, and that could be induced by very weak local coordination, or specific configuration. The presence of such modes would increase the density of states at low-frequency. There is no evidence for their presence in the simulations of [50].

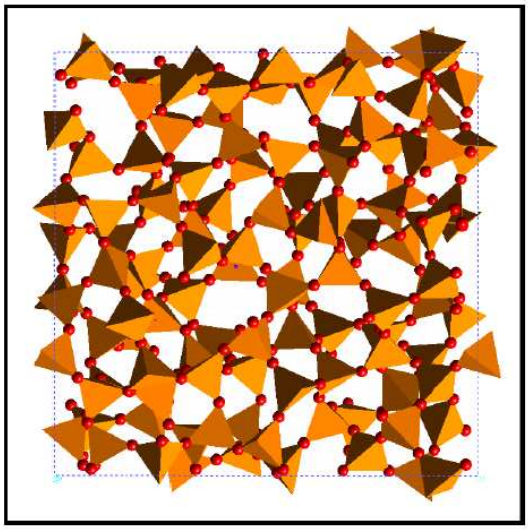

Fig. 4.3: Rigid unit modes model applied to silica. Trachenko et al.[68].

4.3 Application to tetrahedral networks

The nature of slow vibrations -and the possible presence of two-level systems- has been much studied in silica, one of the main glass-forming system. In this glass (or more generally aluminosilicates) the forces within the tetrahedra SiO_4 are much stronger than the forces that act between them [69]: it is easier to rotate two linked tetrahedra than to distort one tetrahedron [8]. This suggests to model such glass as an assembly of linked tetrahedra loosely connected at corners: this is the "rigid unit modes" model [70]. In this model the tetrahedra are characterized by a unique parameter, a stiffness k [9]. Recently this model was used to study the vibrations of silica [71]. The authors first generate realistic configurations of SiO_2 at different pressures using molecular dynamics simulations. At low pressure, they obtain a perfect tetrahedral network. When the pressure becomes large, the coordination of the system increases with the formation of 5-fold defects. Once these microscopic configurations are obtained, the rigid unit model is used and the system is modeled as an assembly of rigid tetrahedra, see Fig.(4.3). Then, the density of states of such network is computed. The results are shown

[8] For example the bending energy of Si-o-Si is roughly 10 times smaller than the stretching of the contact Si-o [72].

[9] In fact the rigidity of a tetrahedron induced by the covalent bonds should be characterized by 3 parameters corresponding to different deformations of the tetrahedron. If these parameters are of similar magnitude, as one expects for example for silica, this does not change qualitatively the results discussed here.

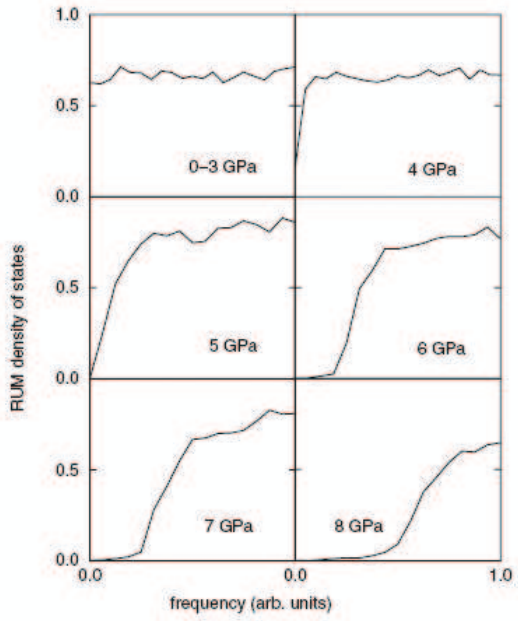

Fig. 4.4: Density of rigid unit modes for silicate at different pressure. Trachenko et al.[71].

in Fig.(4.4). One can note the obvious similarity with the density of states near jamming of Fig.(1.3). We argue that the cause is identical, and that the excess-modes correspond to the anomalous modes made from the soft modes, rather than to one-dimensional modes as proposed in [71]. Indeed, a tetrahedral network is isostatic, see e.g. [68]. The counting of degrees of freedom can be made as follows: on the one hand each tetrahedron has 6 degrees of freedom (3 rotations and 3 translations). On the other hand, the 4 corners of a tetrahedron bring each 3 constraints shared by 2 tetrahedra, leading to 6 constraints per tetrahedron. Thus the system is isostatic. When the pressure increases the coordination increases too, leading to the erosion of the plateau in the density of states discussed earlier in this Chapter. Finally, these predictions of the density of states fail to describe silica glass vibrations at low frequencies, where one cannot neglect the weaker interactions anymore nor the role of the initial stress that we discuss in the next Chapter. In Chapter 8 we evaluate the effect of the weaker interactions. This allows us to propose an explanation for the nature of the boson peak in such glasses.

5. Effect of the initial stress on vibrations

In this section we describe how the above simple description of $D(\omega)$ is affected by the presence of applied stress. In general when a system of particles at equilibrium is formed, there are forces between interacting particles. For harmonic soft spheres it leads to a non-vanishing first term in Eq.(2.9) that becomes:

$$\delta E = \{\frac{1}{4}\sum_{\langle ij\rangle}(r_{ij}^{eq}-1)[(\delta\vec{R}_j-\delta\vec{R}_i)^{\perp}]^2\}+\frac{1}{2}\sum_{\langle ij\rangle}[(\delta\vec{R}_j-\delta\vec{R}_i).\vec{n}_{ij}]^2 \qquad (5.1)$$

where we used $r_{ij}\approx 1$. This term in bracket is (a) negative for repulsive particles (b) proportional to the transverse relative displacement between particle in contact (c) scales as the pressure p, and is therefore vanishing at the jamming transition. The full dynamical matrix $\mathcal{D}$ can be written:

$$\mathcal{D}=\mathcal{M}+\mathcal{M}' \qquad (5.2)$$

where $\mathcal{M}'$ is written in tensorial notation in footnote [10]. The spectrum of $\mathcal{D}$ has *a priori* no simple relation with the spectrum of $\mathcal{M}$. Because $\mathcal{M}'$ is much smaller than $\mathcal{M}$ near the transition, one can successfully use perturbation theory for the bulk part of the normal modes of $\mathcal{M}$. Nevertheless perturbation theory fails at very low frequency, which is of most interest. In this region the spectrum of $\mathcal{M}$ contains the plane waves and the anomalous modes. In what follows we estimate the change of frequency induced by the applied stress on these modes. We show that the relative correction to the plane wave frequencies is very small, whereas the frequency of the anomalous modes can be appreciably changed. Finally we show that these considerations lead to a correction of the Maxwell criterion of rigidity.

[10] in three dimension we have ${\mathcal{M}'}_{ij}=-\frac{1-r_{ij}}{2r_{ij}}[\delta_{\langle ij\rangle}(\vec{m}_{ij}\otimes\vec{m}_{ij}+\vec{k}_{ij}\otimes\vec{k}_{ij})+\delta_{i,j}\sum_{<l>}(\vec{m}_{ij}\otimes\vec{m}_{ij}+\vec{k}_{ij}\otimes\vec{k}_{ij})$, where $(\vec{n}_{ij},\vec{m}_{ij},\vec{k}_{ij})$ is an orthonormal basis.

5.1 Applied stress and plane waves

Consider a plane wave of wave vector k. Since the directions $\vec{n}_{ij}$ are random, both the relative longitudinal and the transverse displacements are of the same order: $[(\delta\vec{R}_i - \delta\vec{R}_j)^{\perp}]^2 \sim [(\delta\vec{R}_i - \delta\vec{R}_j).\vec{n}_{ij}]^2 \sim (\delta\vec{R}_i)^2 k^2$. Consequently the relative correction $\Delta E/E$ induced by the applied stress term is very small:

$$\frac{\Delta E}{E} \approx \frac{\frac{1}{2}\sum_{\langle ij\rangle}(r_{ij}^{eq}-1)[(\delta\vec{R}_j - \delta\vec{R}_i)^{\perp}]^2}{\sum_{\langle ij\rangle}[(\delta\vec{R}_j - \delta\vec{R}_i)\cdot\vec{n}_{ij}]^2} \tag{5.3}$$

since $r_{ij}^{eq}-1$ is proportional to the pressure p, while the others factors remain constant as $p \to 0$, $\frac{\Delta E}{E} \sim p$, and is thus arbitrarily small near the jamming threshold [11]. Note that when the pressure is high, this effect is non negligible. In particular elastic instabilities can occur, and can be responsible for conformational changes, see [73] for such examples in silica crystals.

5.2 Applied stress and anomalous modes

For anomalous modes the situation is very different: we expect the transverse relative displacements to be much larger than the longitudinal ones. Indeed soft modes were built by imposing zero longitudinal terms, but there were no constraints on the transverse terms. These are the degrees of freedom that generate the large number of soft modes. The simplest assumption is that the relative transverse displacements are of the order of the displacements themselves, that is $\sum_{\langle ij\rangle}[(\delta\vec{R}_j - \delta\vec{R}_i)^{\perp}]^2 \sim \sum_i \delta\vec{R}_i^{\,2} = 1$ for the anomalous modes that appear above ω^*. This estimate can be checked numerically for an isostatic system where the sum of the relative transverse terms is computed for all ω. The sum converges to a constant when $\omega \to 0$ as assumed, see Fig.(5.1).

We can estimate the scaling of the correction in the energy ΔE induced by the stress term on the anomalous modes:

[11] In disordered systems the acoustic modes are not exact plane waves, see e.g. the recent simulations in Lennard-Jones systems [6, 7, 8, 9]. As we discuss below, for transverse plane waves the energy is reduced by a factor δz. Therefore the relative correction of energy induced by the applied pressure is of the order of $\frac{p}{\delta z}$, rather than p. We have, as shown, $\frac{p}{\delta z} \ll 1$, so that we still expect the correction to be small near the jamming threshold. In principle non-affine displacements could have other interesting effects, such as an increase of the transverse terms amplitude. If so, the effect of applied stress on acoustic modes would be enhanced.

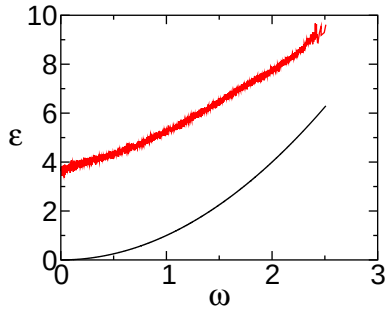

Fig. 5.1: Sum of the transverse terms (red curve) $\epsilon = \frac{1}{2}\sum_{\langle ij\rangle}[(\delta\vec{R}_j - \delta\vec{R}_i)^{\perp}]^2$ and longitudinal terms (black curve) $\epsilon = \frac{1}{2}\sum_{\langle ij\rangle}[(\delta\vec{R}_j - \delta\vec{R}_i)\cdot\vec{n}_{ij}]^2$ for each mode of frequency ω at the jamming threshold in three dimensions. The longitudinal term is equal to the energy of the modes and vanish quadratically at 0 frequency. The transverse term converges toward a constant different from 0.

$$\Delta E \sim -\sum_{\langle ij\rangle}(1 - r_{ij}^{eq})[(\delta\vec{R}_j - \delta\vec{R}_i)^{\perp}]^2 \sim -p \tag{5.4}$$

which is an *absolute* correction, which can be non-negligible in comparison with the energy E.

5.3 Onset of appearance of the anomalous modes

We can now estimate the lowest frequency of the anomalous modes. The modes that appear at ω^* in the relaxed springs system have an energy lowered by an amount of order $-p$ in the original system. Applying the variational theorem of the last section to the collection of slow modes near ω^* indicates that there must be slow normal modes with a lower energy. That is, the frequency ω_{AM} at which anomalous modes appear verifies:

$$\omega_{AM} \leq [(\omega^*)^2 - A_2 p]^{\frac{1}{2}} \equiv [A_1\delta z^2 - A_2 p]^{\frac{1}{2}} \tag{5.5}$$

where A_1 and A_2 are two positive constants. This indicates that the important parameters of the low frequency excitations are coordination and stress.

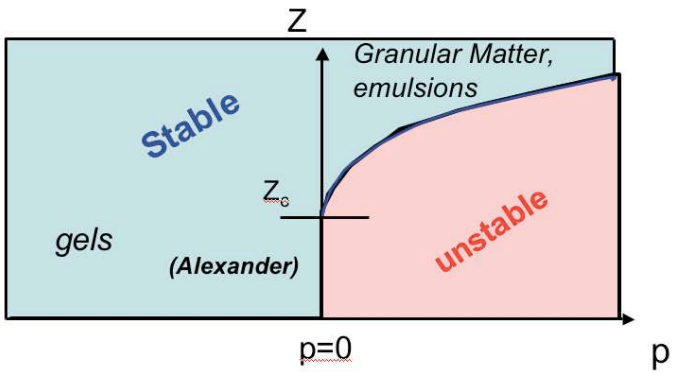

Fig. 5.2: Phase diagram of rigidity in terms of coordination and pressure. When $p > 0$, the line separating the stable and unstable regions is defined by Eq.(5.6). When the pressure is negative, any connected system can be rigid, as it is the case for gels.

5.4 Extended Maxwell criterion

From this estimation we can readily obtain a relation among coordination and pressure that guarantees the stability of a system. There should be no negative frequencies in a stable system, therefore $\omega_{AM} > 0$. Thus in an harmonic system the right hand side of Eq.(5.5) must be positive:

$$\delta z \geq C_0 p^{\frac{1}{2}} \equiv \delta z_{min} \tag{5.6}$$

where C_0 is a constant. This inequality, which must hold for any spatial dimension, indicates how a system must be connected to counterbalance the destabilizing effect of the pressure. A phase diagram of rigidity is represented in Fig.(5.2). When $p = 0$, the minimal coordination z_c corresponds to the isostatic state: this is the Maxwell criterion. As we said earlier for spherical particles $z_c = 2d$. As we discuss later $z_c = d + 1$ for particles with friction. When $p > 0$, Eq.(5.6) delimits the region of rigid systems: for example granular matter or emulsions lie above this line. When $p < 0$, even systems far less connected than what the Maxwell criterion requires are rigid [36], as we discussed in Chapter 2. These systems contain many soft modes as defined in Eq.(2.3), but there are all stabilized by the positive bracketed term of Eq.(2.9). Note that a similar phase diagram, with the same singularity of δz but a different z_c, was recently obtained by a mean field approach [74].

Finally the relation (5.6) is verified in the simulations of [50] where $\delta z \sim p^{\frac{1}{2}}$. In the next Chapter we justify why the inequality (5.6) is in fact an equality in the system of [50] and discuss what determines the microscopic

structure of this system.

6. Microscopic structure and marginal stability

In the previous section we studied how the low-frequency vibrational properties were related to the microscopic structure. The applied pressure has two antagonist effects: on the one hand, it increases the coordination number, which stabilizes the system and increase the frequency of appearance of the anomalous modes. On the other hand, the applied pressure appears in the expansion of the energy and lowers the frequency of the anomalous modes. In what follows we study the relative amplitude of these two effects, or equivalently, where such amorphous solids are located in the (z, p) plane of Fig.(5.2).

Since these systems are out-of-equilibrium, their microscopic structures depend on the system history. As we shall see below, the preparation of the system of [50] leads to *marginally stable* systems at any ϕ. The two antagonists effects of the pressure compensate [12], which leads to $\omega^* \gg \omega_{AM}$, as it appears in Fig.(4.2), and $\delta z \sim p^{\frac{1}{2}}$. In what follows we propose a simple argument to justify such a behavior. We discuss in particular (i) the dynamic that takes place when a liquid of repulsive spheres is hyperquenched and (ii) the decompression of a jammed solid at zero temperature. This will also enable us to discuss the surprising geometrical property of the random close packing evoked in the introduction: there is a divergence in the pair correlation function $g(r) \sim 1/\sqrt{r-1}$ at close contact. We propose that this divergence is a vestige of the marginal stability that occurs at higher packing fraction.

6.1 Infinite quench

The simulations of [50] show $\delta z \sim p^{\frac{1}{2}}$, thus saturating the bound of Eq. (5.6), so that there are excess modes extending to frequencies much less than ω^*.

[12] Assuming an exact compensation of these two terms lead to $\omega_{AM} = 0$ in an infinite size system. In Fig.(4.2) ω_{SM} is slightly different from zero, as one would expect for a finite size system.

We start by furnishing an example of dynamics that lead to such a situation. Consider an initial condition where forces are roughly balanced on every particle, but such that the inequality (5.6) is not satisfied. Consequently, this system is not stable: infinitesimal fluctuations make the system relax with the collapse of unstable modes. Such dynamics was described by Alexander in [36] as structural buckling events: they are are induced by a positive stress as for the buckling of a rod, but take place in the bulk of an amorphous solid. These events a priori create both new contacts and decrease the pressure. When the bound of (5.6) is reached, there are no more unstable modes. If the temperature is zero, the dynamics stops. Consequently one obtains a system where Eq.(5.6) is an *equality*, therefore (i) this system is weakly connected (ii) $\omega_{AM} = 0$, there are anomalous modes much different from plane waves extending to zero frequency. A similar argument is present in [74].

In the simulations of Ref [50] the relaxation procedes as follows. The system is initially in equilibrium at a high temperature. Then it is hyperquenched to zero temperature. At short time scales the dynamics that follows is dominated by the relaxation of the stable, high frequency modes. The main effect is to restore approximately force balance on every particle. At this point, if inequality (5.6) is satisfied, the dynamics stops. If it is not satisfied, we are in the situation of buckling described above. The pressure and co-ordination number continue to change until the last unstable mode has been stabilized. At this point the bound of Eq. (20) is marginally satisfied, and there is no driving force for further relaxation.

It is interesting to discuss further which situations lead to marginal stability. We expect that the situation of marginal stability that follows an infinite cooling rate takes place for a domain of the parameters of initial conditions (ϕ, T), located at high temperature and low density. This domain could end at a finite ϕ even when temperature is infinite, as it does in some related systems. This was shown in simulations and theoretical work on Euclidian random matrix [42] were most of the unstable modes vanish beyond a finite ϕ even when $T = \infty$. When the cooling rate is finite, as in experiments, we expect that the relaxation does not stop when all the modes are stable, but that there are activated events that lead to further collapses of anomalous modes. These events a priori increase the connectivity and decrease the pressure beyond the bound of Eq.(5.6), leading to $\omega_{AM} > 0$. In agreement with this idea, hyperquenched mineral glasses show a much stronger excess of modes[76] in comparison with normally cooled glasses, and annealed polymeric glasses expectedly show a smaller excess of modes [77].

6.2 Decompression

The data of [50] were obtained by gradually decreasing the pressure from the above initial state of zero temperature and nonzero pressure. When pressure is lowered, it is observed that the system remains marginally stable. Here we propose a qualitative interpretation of these findings in the case of harmonic particles [13]. In [50], the decompression is obtained by discrete steps. At each step, the radii of each particle is reduced by a small amount ϵ, while the center of the particles are kept fixed. This corresponds to an affine decompression. This new configuration is not at equilibrium in general. Then the particles are let to relax. The affine decompression has two effects: on the one hand it causes some contacts to open, on the other hand it reduces the pressure. The opening of these contacts tends to destabilize normal modes and reduce their frequencies, while the reduction in pressure tends to stabilize them. As we argue below, the destabilizing effect dominates. Thus, the affine decompression leads the system into the unstable region. Therefore we expect that when the particles relax, one recovers the dynamics that follows an infinite quench: modes buckle. As for the infinite quench of temperature discussed above, buckling should occur as long as the relaxation of the stable normal modes, which is faster than the collapse of unstable modes, does not bring back the system into the stable region. If so, the buckling increases the contact number and decreases the pressure until marginal stability is achieved, so that the inequality of Eq. (5.6) is marginally satisfied as the pressure decreases, as observed in the simulations of [50].

In what follows we justify the claim that the destabilizing effect of the opening of the contacts dominate the effect of the pressure reduction. When the particles radii decrease by an amount ϵ, a certain fraction $e \sim g(1)\epsilon$ of contacts opens, where $g(r)$ denotes the radial distribution function. For harmonic particles we expect $g(1) \sim (\phi - \phi_c)^{-1}$ [14]. Hence, using that $p \sim (\phi - \phi_c)$ for harmonic particles [50] –as we shall demonstrate in the next Chapter–, one obtains $e \sim \frac{\epsilon}{p}$. On the other hand, the affine decompression lowers the pressure by an amount $\delta p \sim \delta\phi \sim \epsilon$. Thus the system can only

[13] To extend this argument to other potentials, for example Hertzian contacts, one should assume that $g(1)$ scales at least as $(\phi - \phi_c)^{-\frac{1}{2}}$ when the jamming threshold is approached. This would imply that the number of contacts lost when the system is decompressed is large enough to generate buckling. This point is related to the evolution of the force distribution of Hertzian particles near the isostatic point. It is a subtle issue, and we are not aware of any numerical results of this sort.

[14] This is related to well-known empirical facts of the force distribution: one has $P(F)dF \sim g(r)dr$. For harmonic particles $dr \sim dF$ and therefore $P(F) \sim g(r)$. When rescaled by $\sim p$, $P(F)$ converges to a master curve with $P(F/\langle F\rangle = 0) \neq 0$ [50, 51]. This implies that $g(1) \sim p^{-1} \sim (\phi - \phi_c)^{-1}$ [78].

afford to lose a fraction f of contacts while remaining stable: according to Eq.(5.6): $f = \frac{d(\delta z_{min})}{dp}\delta p \sim p^{-\frac{1}{2}}\delta p \sim \frac{\epsilon}{p^{\frac{1}{2}}} \ll e$. Therefore $f \ll e$ as claimed. Hence if an affine reduction of packing fraction ϵ is imposed, far too many contacts open and the system is unstable.

To conclude, it was observed in the simulations of [50] that if the steps ϵ are small, the decompression that takes place in [50] is reversible: cycles of decompression-compression bring the system back to its initial configuration. This empirical fact indicates the absence of discontinuous irreversible events. Thus, the buckling generated by the opening of few contacts when ϵ is small enough does not lead to rearrangements of finite amplitude much larger than ϵ. This indicates that the dynamic of modes collapse increases the coordination by re-closing most of the contacts that open during the affine decompression (whose particles are separated by distance of at most ϵ), and not by forming new contacts. It is reasonable to think that if several cycles of compression/decompression are made, the system will end up to be reversible. Why it is already so at the first decompression is a subtle question that we do not try to justify here.

6.3 $g(r)$ at the random close packing

The probability $g(r)$ of having two particles separated by a vector of length r displays a square root divergence $g(r) \sim (r-1)^{-\frac{1}{2}}$ at the jamming transition. In the introduction we pointed out that this divergence corresponds to the singular increase of the excess coordination δz with the packing fraction, as was noted in [50]. Here justify further this correspondence. We show that, if the decompression is assumed to be adiabatic, the singularity in $g(r)$ is a necessary consequence of the marginal stability that characterizes the decompression. We argue that the pair of particles almost touching at ϕ_c, responsible for the divergence in $g(r)$, are the vestiges of the contacts that were present at higher ϕ to stabilize the system. In order to show this, we first have to count the contacts that open for a given $\phi - \phi_c$. Then we shall estimate the distance between the corresponding pairs of particles at the jamming threshold.

As we discussed in the last section, when the system is decompressed it remains marginally stable: the coordination follows Eq.(5.6), and $z = z_c + \delta z_{min}$. Thus the density of contacts $n(\phi)$ per unit ϕ that open for a given ϕ follows exactly in the large L limit:

$$n(\phi) \sim \frac{d(\delta z_{min})}{d\phi} \sim \frac{1}{(\phi - \phi_c)^{\frac{1}{2}}} \tag{6.1}$$

We now would like to evaluate the distance that separates such particles at the jamming threshold. Let h_ϕ be the random variable that corresponds to the spacing $r-1$ at the jamming threshold between two neighboring particles whose contact opened at a given ϕ. We want to estimate the fluctuations of h_ϕ. If the decompression was purely affine h_ϕ would be single-valued and given by $h_{aff}(\phi)$:

$$h_{aff}(\phi) = \frac{\phi - \phi_c}{d\phi_c} \tag{6.2}$$

As we discussed in the last section the displacements that follow a decompression are not affine. For our present argument we need to evaluate the variance of the distribution of h_ϕ. It is directly related to the variance of the non-affine displacements that appear while decompressing. We expect that such non-affine displacements simply lead to a variance of h_ϕ of order of its average $h_{aff}(\phi)$ [15]. Therefore the probability $P_\phi(h)$ that two particles whose contact opened at a given ϕ are at a distance $1+h$ at the threshold can be written as:

$$P_\phi(h) \equiv \frac{1}{\phi - \phi_c} f(\frac{h}{\phi - \phi_c}) \tag{6.3}$$

where f is a normalized scaling function $\int f(x)dx = 1$. Thus one can compute $g(r)$ by summing over all the contributions of the contacts that opened at $\phi > \phi_c$, as we represent on Fig.(6.1):

$$g(r) \sim \int n(\phi) P_\phi(h = r-1) d\phi \sim \int \frac{1}{(\phi - \phi_c)^{3/2}} f\left(\frac{r-1}{\phi - \phi_c}\right) d\phi \tag{6.4}$$

$$g(r) \sim \frac{1}{(r-1)^{\frac{1}{2}}} \int u^{-\frac{1}{2}} f(u) du \tag{6.5}$$

[15] More generally, if two particles (in contact or not) are at a distance r of order one at a given ϕ, we expect that the fluctuations of the distance that separate them at ϕ_c due to non-affine displacements is of order of the affine increase of their distance $\delta r = r\frac{\phi-\phi_c}{d\phi_c} \approx \frac{\phi-\phi_c}{d\phi_c} = h_{aff}(\phi)$. This comes from the following observation: non-affine displacements are induced by the requirement of stability that creates correlations among particles motions. To evaluate such correlations, consider the typical situation discussed in the last section where particles in contact at a given ϕ have to stay in contact until ϕ_c, instead of spreading apart if the displacement was affine. At ϕ_c the inter-particle distance is 1, instead of a value $r < 1 + h_{aff}(\phi)$. Thus we evaluate the typical departure from a pure affine displacement to be of order $h_{aff}(\phi)$.

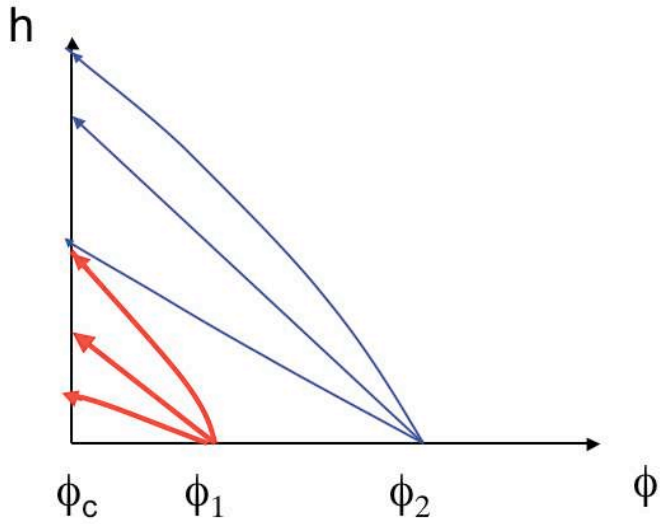

Fig. 6.1: Representation of typical variations of the spacing h between particles whose contacts opened at given $\phi > \phi_c$. The arrows are typical trajectories. Those starting from ϕ_1 are thicker than those starting from ϕ_2, indicating that the density of contacts that open is larger as ϕ decreases toward ϕ_c. $g(r)$ at ϕ_c is computed by summing over all trajectories for all $\phi > \phi_c$.

We do not expect any singularity of $f(u)$ in $u = 0$ [16], and therefore:

$$g(r) \sim \frac{1}{(r-1)^{\frac{1}{2}}} \tag{6.6}$$

6.4 Isostaticity, $g(r)$ and thermodynamics

In our argument above the divergence in $g(r)$ does not appear through a minimization of a free energy or volume, but rather as a consequence of the specific dynamics that took place to reach the jamming threshold. Empirically, it turns out that the divergence in $g(r)$ was observed in systems obtained with different dynamics. It was first noticed in molecular dynamics simulations of elastic spheres with friction, where the jammed states were

[16] We do not exclude that there is a finite probability p_0 for particles that lost their contact at a given ϕ_0 to recover it at a ϕ_1 such that $\phi_c < \phi_1 < \phi_0$. In our formalism such eventuality will not create a divergence in $f(u)$. The incoming flux of contact would be compensated by a larger rate of contact opening, and the expression (6.1) will be corrected by a factor $\frac{1}{1-p_0}$.

obtained using inelastic collisions [53]. It was also recently observed — with possibly a slight difference in the exponent— in random close packing of hard spheres obtained by increasing continuously the packing fraction from the liquid state [79]. In this last example thermodynamic equilibrium is maintained until fairly high packing fractions, corresponding to typical inter-particle spacing much smaller than the particles radii [17]. On the other hand in this hard sphere system the range of validity of Eq.(6.6) extends to rather large distances, say $r - 1 \approx 0.3$. Thus, in this system the divergence of $g(r)$ at ϕ_c cannot be the vestige of an interaction that would have taken place between pairs of neighbor particles at lower ϕ, as the system has a finite relaxation time and neighbor particles can move arbitrary far from each other. Instead, this suggests that the divergence of $g(r)$ obtained in [79] is related to thermodynamic properties of a hard sphere liquid near random close packing.

In what follows we discuss why the properties of isostaticity and divergence of $g(r)$, which were introduced in relation with the mechanical stability of the solid phase, may also be connected to the thermodynamic of the liquid phase of hard spheres (see also Chapter 9). Our argument is in two steps. We first argue that the isostaticity and the divergence in $g(r)$ have an interpretation in terms of density phase space of hard sphere packing. We shall show that if an assembly of hard spheres is (i) sub-isostatic, that is with a coordination smaller than $2d$, or (ii) isostatic with a singularity of $g(r)$ in $r = 1$ weaker than Eq.(6.6), then it is not a "good" local maximum of density: one can build close configurations which are denser. In a second step, we simply argue that in the liquid phase, the system prefers to lie in configurations close to the denser packing, which is favorable for entropic reasons. Finally this suggests that when ϕ increases toward ϕ_c, the system becomes isostatic and display at least a square root singularity in $g(r)$ at $r = 1$.

Consider a sub-isostatic configuration of hard sphere in a box. As we discussed in earlier Chapters, this system is not rigid. Therefore, if an infinitesimal pressure is applied at the boundary of the box, the system cannot resist to it and yields. Unstable modes collapse until isostaticity is recovered. Consequently, the volume of the box decreases and one obtains a denser configuration of spheres. The same kind of argument can be made if one considers an isostatic configuration of hard spheres, again contained in a box, that do not display the square root singularity of $g(r)$ in $r = 1$. Let assume temporally that the particles are not hard, but soft, and interact for example with harmonic interactions. Such system is again unstable toward

[17] For mono-disperse spheres thermodynamic equilibrium can be obtained below typically $\phi \approx 0.58$, which corresponds to an inter-particle spacing $h \approx 0.03$

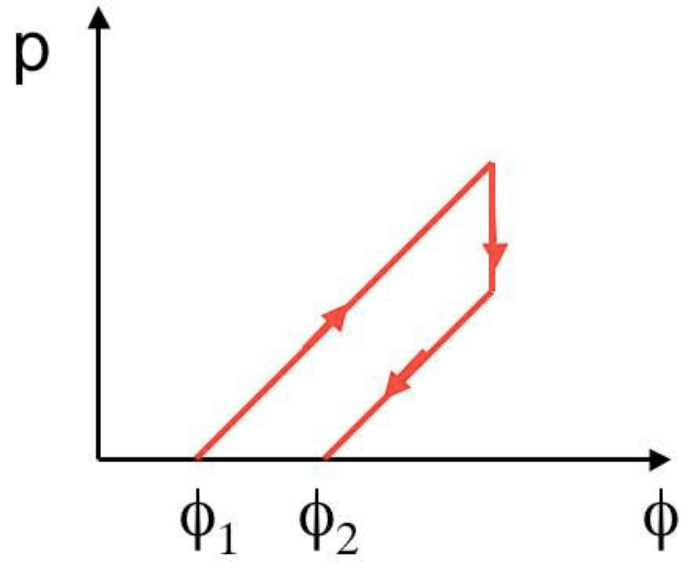

Fig. 6.2: Cycle of compression-buckling-decompression starting from a packing fraction ϕ_1 and ending at a larger packing fraction ϕ_2.

compression: an affine compression of strain ϵ increases the coordination by an amount that scales as $\delta z \sim \int_1^{1+\epsilon} g(r)dr$. If $g(r)$ does not display the square root singularity in $r = 1$, one finds $\delta z \ll \delta z_{min} \sim p^{\frac{1}{2}} \sim \epsilon^{\frac{1}{2}}$. Thus unstable modes collapse until the system is enough coordinated, which lowers the pressure. Then if the pressure is brought back to 0 adiabatically, one obtains a new isostatic state of higher density, as sketched in Fig.(6.2).

Now we argue that a liquid of hard spheres lies near the dense jammed states. More precisely, we define the jammed state corresponding to a liquid configuration as the solid obtained by hyperquenching the liquid, that is by increasing rapidly the radii of the particles until a solid is formed. We argue that the densest jammed state are favored, as they offer a larger free volume per particle in the liquid configuration. Consider for example two jammed states of packing fraction $\phi_a < \phi_b$, and a liquid at packing fraction $\phi < \phi_a$. If the liquid lies near the jammed state a the free volume per particles varies as $\phi_a - \phi$. Thus if p_a is the probability for the liquid to have a as the jammed state, one can estimate $p_a/p_b \sim (\frac{\phi_a - \phi}{\phi_b - \phi})^N$. Thus the densest jammed states are favored as the packing fraction increases. For a quantitative treatment of the liquid equilibrium, one should enumerate all the possible jammed states that are accessible to the dynamics —as there are configurations such as the crystal which are not visited in a reasonable time— and compute their structure. A similar approach was followed recently in [80]. Our qualita-

tive argument suggests that the isostatic states with the divergence in $g(r)$ evoked above are good candidates to dominate the dynamics near the glass transition, since they are (i) good (in the sense stated above) local maxima of density and (ii) plausibly numerous, as we exhibited a simple way to create them. If so, $g(r)$ in a random close packing obtained from the liquid phase must display a square root divergence.

7. Elastic response near the jamming threshold

In this Chapter we first derive a few properties of the response to a local strain, such as the deformation of one contact, near the jamming threshold. We show that at the isostatic limit, such a response spreads out uniformly in the entire system. This implies that the soft modes of an isostatic state obtained by cutting one contact are extended, as assumed in Chapter 3. When the coordination increases, we show that the energy cost induced by such a local strain grow as δz.

This suggests that the shear modulus also scales as δz, as it is observed numerically. We confirm this behavior by computing the response of the system to a global compression and a global shear. An assembly of repulsive short-range particles near jamming behaves as a gel in the following sense: the shear modulus is much smaller than the bulk modulus, as it was observed numerically [50, 81]. In a gel, the response to compression is the one of liquid of monomers, whereas the response to a shear is related to the stretching of the polymer chains. The cause for the discrepancy between these two moduli is different near jamming. If the responses to compression and shear were purely affine, then the two moduli would scale in the same way. As we shall show the bulk modulus is simply what one would expect from an affine response of the system. In the case of a shear, the non-affine displacements lower the energy dramatically, which even vanishes at the transition. This is possible because the low-coordination allows particles to rearrange without much energy cost. Thus this system is a useful tool to study non-affine displacements, that might play an important role to describe material failure [82], and that were observed in various systems such as granular matter or biological materials, and in simulations of glasses and foams, see e.g. [83] for references. We discuss very tentatively the possible length scales that characterize such displacements.

7.1 Formalism

In this section we study the relation between forces and soft modes. We show that the equation of force balance is the dual of the equation that defines the soft modes. This will be used in the next section to compute the response to a strain.

7.1.1 Force propagation

Many of the properties discussed below concern the response of the system to external forces. It proves convenient to consider our system under the influence of an arbitrary set of forces $\vec{F}_i$ acting on all particles i. At equilibrium the sum of the forces on each particle i is null:

$$\sum_{\langle j \rangle} f_{ij} \vec{n}_{ij} = \vec{F}_i \tag{7.1}$$

where f_{ij} is the compression in the contact $< ij >$, the sum is on the particles in contact with particle i, and $\vec{F}_i$ the external force applied on i. In term of sign convention we recall that $\vec{n}_{ij}$ is the unit vector going from i to j. This linear equation can be written:

$$\mathcal{T}|\mathbf{f}\rangle = |\mathbf{F}\rangle \tag{7.2}$$

$|\mathbf{f}\rangle$ is the vector of contact tensions and has N_c components. $\mathbf{F}$ is the vector of external forces. Its dimension is $(Nd - d(d+1)/2)$. Indeed there are d degrees of freedom for the external force on each particle, which brings the term Nd. The term $d(d+1)/2$ corresponds to the constraints on the total torques and forces that must be zero at equilibrium (in what follows our notations shall be lower-case for the contact space of dimension N_c, and upper-case for the particles positions space of dimension $Nd - d(d+1)/2$). Therefore $\mathcal{T}$ is an $Nd - d(d+1)/2 \times N_c$ matrix.

In the case where $N_c > Nd - d(d+1)/2$ equation (7.2) is not sufficient to fix the contact tensions. One must apply elasticity theory to compute the contact forces for a set of external forces. This is done by writing the relation between forces and distances between particles in contact (which depends on the interaction potential), and by relating the distance among particles to their position. This brings the necessary $N_c - Nd - d(d+1)/2$ extra-constraints. When $N_c < Nd - d(d+1)/2$, there is in general no solution for a given force field, and the system yields under a generic external force field. For such a system no contact forces $\mathbf{f}$ can satisfy (7.2) unless $\mathbf{F}$ is restricted. In the last case, when $N_c = Nd - \frac{d(d+1)}{2}$, the system is called

"isostatic", and there are just enough contacts to equilibrate any external force field. Note that in this case Eq.(7.2), which does not depend on the interaction potential, is sufficient to determine the contact forces.

7.1.2 Duality between force propagation and soft modes

As we discussed in earlier Chapters, the static equilibrium of the system can be expressed in geometric terms, rather than in terms of the forces as in Eq.(7.2). The change of distance to the first order between particles for a given set of displacements $\delta\vec{R}_i$ is:

$$(\delta\vec{R}_i - \delta\vec{R}_j).\vec{n}_{ij} = \delta r_{ij} \tag{7.3}$$

It can be written as:

$$\mathcal{S}|\delta\mathbf{R}\rangle = |\delta\mathbf{r}\rangle \tag{7.4}$$

where $\delta\mathbf{r} \equiv \{\delta r_{ij}\}$ is the set of distance changes for all contacts. If one removes the global translations or rotations from the displacement fields, which obviously do not change any inter-particle distance, $\mathcal{S}$ becomes an $(Nd - d(d+1)/2)$ by N_c matrix. Its kernel is the space of soft modes we introduced in Chapter 2. When $N_c < Nd - d(d+1)/2$, the displacements are undetermined for a given change of inter-particles distances.

Now we establish the connection among soft modes and forces, which is also present in a similar form in [58, 65]. At equilibrium for any displacement, the net work done by external forces and the contact forces is zero: this assures the stability of the system. Therefore for any acceptable equilibrium force field:

$$\sum_i \delta\vec{R}_i \cdot \vec{F}_i - \sum_{ij} \delta r_{ij} f_{ij} = \langle\delta\mathbf{R}|\mathbf{F}\rangle - \langle\delta\mathbf{r}\cdot\mathbf{f}\rangle = 0 \tag{7.5}$$

For a soft mode, $\delta r_{ij} = 0$ and we are left with:

$$\langle\delta\mathbf{R}|\mathbf{F}\rangle = \sum_i \delta\vec{R}_i \cdot \vec{F}_i = 0 \tag{7.6}$$

The soft modes are equivalent to the constraints on the force field one would have obtained by solving Eq.(7.1). Each soft mode represents a direction of fragility of the system, and the external forces must be orthogonal to them to avoid yielding. Furthermore, applying the definitions of $\mathcal{S}$ and $\mathcal{T}$ in (7.5) we have:

$$\langle\mathbf{f}|\mathcal{S}|\delta\mathbf{R}\rangle = \langle\delta\mathbf{R}|\mathcal{T}|\mathbf{f}\rangle \tag{7.7}$$

therefore, introducing the transpose notation, we obtain:

$$\mathcal{S} = \mathcal{T}^t \tag{7.8}$$

7.1.3 Relation with the Dynamical matrix

In this Chapter we shall neglect the initial stress term in the energy expansion, as we did in Chapters 3 and 4. As we discussed in Chapter 5, we expect that this term affects only weakly the plane waves. Thus it should lead to negligible corrections in the computations of the responses at zero wave vector such as a shear or a compression. In the harmonic case we discuss, we have then:

$$\delta E = \sum_{<ij>} \frac{1}{2}\delta r_{ij}^2 = \frac{1}{2}\langle \mathcal{S}\delta\mathbf{R}|\mathcal{S}\delta\mathbf{R}\rangle \tag{7.9}$$

Therefore one finds for the dynamical matrix defined in Eq.(2.2):

$$\mathcal{M} = \mathcal{S}^t\mathcal{S} \tag{7.10}$$

7.2 Relation between the response to a strain and forces

The response to an imposed strain is given by the minimization of the energy with respect to the displacement fields. In this section, using simple linear algebra and the duality between force and soft modes studied above, we show how this minimization can be written as a sum over the vectors in the contact force space that satisfy force balance. This will allow in the next sections to (i) derive exact results of the response to a local strain (ii) derive the bulk and shear moduli assuming well-known empirical fact of force properties, that we recall at the end of this section.

As we discussed in earlier Chapters, in the approximation where the initial stress is neglected, the system is equivalent to a set of point particles interacting with springs. In order to study the elastic behavior of such system, it turns out to be convenient to consider the responses that follow arbitrary changes of rest length of these springs. This is in fact equivalent to imposing dipoles of force. Consider for example two particles i and j in contact, and increase the rest length of their spring by an infinitesimal amount y_{ij}. It is equivalent to impose a dipole of force where opposite external forces are imposed on i and j with $\vec{F}_j = -\vec{F}_i = y_{ij}\vec{n}_{ij}$. As we shall see, the response to shear or compression can also be easily expressed in terms of changes of rest length.

We impose an infinitesimal change of rest length on every contact $\mathbf{y} = \{y_{ij}\}$. Following Eq.(7.9) the energy and the displacement field are given by the minimization of:

$$\delta E = \frac{1}{2} \min_{\{\delta \mathbf{R}\}} \langle \mathcal{S}\delta\mathbf{R} - \mathbf{y} | \mathcal{S}\delta\mathbf{R} - \mathbf{y} \rangle \tag{7.11}$$

Obviously if $\mathcal{S}$ was spanning its image space, we would have $\delta E = 0$: one could always find a displacement $|\delta\mathbf{R}\rangle$ that leads to a change of distances between particles in contact exactly equal to $|\mathbf{y}\rangle$. As we said, $\mathcal{S}$ is a $N_c \times Nd - d(d+1)/2$ matrix. If $N_c < Nd - d(d+1)/2$, $\mathcal{S}$ indeed spans its image space, and the energy associated with any strain $|\mathbf{y}\rangle$ is zero: the system is floppy. In the other case, if $N_c > Nd - d(d+1)/2$, there are $N_c - Nd - d(d+1)/2 \equiv N\delta z/2$ relations of dependency among the columns of $\mathcal{S}$. One can choose a basis of $N\delta z/2$ vectors $|\mathbf{a}^p\rangle$, with $1 \leq p \leq N\delta z/2$, in the space of $|\delta\mathbf{r}\rangle$ such that:

$$\langle \mathbf{a}^p | \mathcal{S} = 0 \tag{7.12}$$

The $|\mathbf{a}^p\rangle$ are orthogonal to all the vectors $\mathcal{S}|\delta\mathbf{R}\rangle$, for any displacement field $|\delta\mathbf{R}\rangle$. Transposing this relation we have:

$$\mathcal{T}|\mathbf{a}^p\rangle = 0 \tag{7.13}$$

which indicates that all the vectors in the space of the $|\mathbf{a}^p\rangle$ satisfy force balance without external force (7.1), but no others. The $|\mathbf{a}^p\rangle$ live in the contact-force space, and henceforth we shall denote them $|\mathbf{a}^p\rangle \equiv |\mathbf{f}^p\rangle = \{f_{ij}^p\}$. In the following we consider an orthogonal unit basis:

$$\langle \mathbf{f}^p | \mathbf{f}^{p'} \rangle \equiv \sum_{ij} f_{ij}^p f_{ij}^{p'} = \delta_{pp'} \tag{7.14}$$

We can decompose any $|\mathbf{y}\rangle$ as:

$$\mathbf{y} = \mathbf{y}^\perp + \sum_{p=1\ldots N\frac{\delta z}{2}} \langle \mathbf{f}^p | \mathbf{y} \rangle \mathbf{f}^p \tag{7.15}$$

$\mathbf{y}^\perp$, the part of $|\mathbf{y}\rangle$ orthogonal to the $|\mathbf{f}^p\rangle$, is spanned by the matrix $\mathcal{S}$, and therefore does not contribute to the energy when the minimization of Eq.(7.11) is performed. In other words, there is a displacement field which leads to a strain $\mathbf{y}^\perp$. The energy that results of the minimization of Eq.(7.11) is then simply given by the distance square between $|\mathbf{y}\rangle$ and the image vector space of $\mathcal{S}$, that is $||\mathbf{y} - \mathbf{y}^\perp||^2$ or:

$$\delta E = \frac{1}{2} \sum_{p=1,..\frac{N\delta z}{2}} \langle \mathbf{f}^p | \mathbf{y} \rangle^2 \tag{7.16}$$

Furthermore the response to such strain condensates an energy δE_{ij} in the contact $< ij >$ that follows:

$$\delta E_{ij} = \frac{1}{2}(y_{ij} - y_{ij}^{\perp})^2 = [\sum_{p=1,..\frac{N\delta z}{2}} \langle \mathbf{f}^p | \mathbf{y} \rangle f_{ij}^p]^2 \tag{7.17}$$

Expressions (7.16) and (7.17) furnish explicit solutions for the minimization of Eq.(7.11). They do not give access to the displacement field $\delta\mathbf{R}$ that follows a given strain (that can be computed in principle by inverting $\mathcal{S}$), but to the energy condensed globally or in any contact. From these expressions we shall derive a few exact properties of the response to a localized strain or force dipole. Furthermore, they relate the response to a strain to properties of forces, which are well studied empirically. Making simple assumptions on these contact force fields, we shall derive the scaling of the elastic moduli.

7.3 Response to a local perturbation

We impose a local strain corresponding to the compression or the stretching of one contact. We implement it by increasing the separation r_{12} of a single contact 12 chosen to be at the origin by an amount ϵ. This is equivalent to impose a force dipole $\vec{F_1} = -\vec{F_2} = \epsilon \vec{n}_{12}$. We have:

$$y_{ij} = 0 \text{ if } ij \neq 12 \tag{7.18}$$

$$y_{12} = \epsilon \tag{7.19}$$

It is now straightforward to compute the energy cost of local strain using Eq.(7.16). One finds:

$$\delta E = \frac{1}{2} \sum_{l=1,..N\frac{\delta z}{2}} (f_{12}^l \epsilon)^2 \tag{7.20}$$

If one sums this expression for every possible contacts ij and use the normalization of the vectors of force contact, one finds exactly $N\frac{\delta z}{4}\epsilon^2$. Therefore we have the following exact result for the average energy induced by such deformation:

$$\langle \delta E \rangle = \frac{\delta z}{2z}\epsilon^2 \tag{7.21}$$

This results can be extended to the case where several contacts are deformed. For example, if a particle swells, all its contacts are compressed by the same factor ϵ, and one finds that the energy still goes as δz. In a isotropic continuous elastic medium with a weak shear modulus G, the energy resulting from the swelling of a small sphere goes as G. Thus we may infer from Eq.(7.21)

that $G \sim \delta z$. This scaling is observed numerically [50]. In the next section we shall justify further this scaling law by computing the response to a global shear.

Eq.(7.21) shows that when the jamming threshold is approached, deforming a contact (or imposing a force dipole) becomes softer and softer. Eventually, when $\delta z \to 0$, the deformation of the contact becomes totally soft. The response corresponds then exactly to the soft modes that we described in Chapter 3 that appear when a contact (here the contact 12) is removed in an isostatic system. A fundamental question was the extension of such modes, that was assumed to be over the whole system at that point. In what follows we show that this is indeed the case.

When the jamming threshold is approached, the coordination diminishes until there is eventually only one term left in the sum of Eq.(7.20). At that point there is one more contact than in an isostatic state [18] and $\delta z = \frac{2}{N}$. Eq.(7.21) indicates that the energy resulting from the deformation of one contact is of order $1/N$, which vanishes as the system size increases. The deformation becomes soft. We can use Eq.(7.17), whose sum contains only one force field, to obtain the spatial repartition of the energy:

$$\delta E_{ij} = \frac{\epsilon^2}{2}(f^1_{12} f^1_{ij})^2 \tag{7.22}$$

$|\mathbf{f^1}\rangle$ is the physical set of contact forces that support the system. These contact forces are well studied numerically and experimentally, and we shall discuss more their properties in the next paragraph. At large distance where the contact force are de-correlated $\langle (f^1_{12} f^1_{ij})^2 \rangle = \langle (f^1_{ij})^2 \rangle^2 > 0$: the energy condensed in the contact ij does not vanish when the distance between ij and 12 diverges. If the displacements that follow such a perturbation were to vanish with the distance from the source, it would be so for the energies condensed in the contacts. As it is not the case, these displacements spread out in the entire system.

[18] At the jamming threshold, the system is in fact not exactly isostatic: it has one extra contact. It must be the case as there must be a non-zero contact force field to pin the particles. The counting of degrees of freedom of Chapter 2, $N_c = Nd - d(d+1)/2$, is in fact incomplete: it must take into account the degree of freedom corresponding to the box size. This subtle point, which does not affect the result of the previous Chapters, is trivial in one dimension, as N particles pinned on a ring have N contacts, and not $N-1$. This was verified numerically in the simulations of [50] in 2 and 3 dimensions as removing m contacts at the threshold creates only $m-1$ soft modes instead of m.

7.4 Elastic moduli

7.4.1 Spatial properties of the force fields $|\mathbf{f}^p\rangle$

In this paragraph we discuss the properties of the contact force fields that we shall use to derive the scaling of the bulk and the shear moduli. Only one vector of the vector space of the force fields $|\mathbf{f}^p\rangle$ solutions of Eq.(7.2) without external force is the real set of contact forces that supports the system. As we discussed at the beginning of this Chapter, it can be computed by solving the whole elastic problem. The solution of such problem is unique. This vector is denoted $|\mathbf{f}^1\rangle$. The rest of the basis $|\mathbf{f}^p\rangle$ with $p \neq 1$ are also solutions of Eq. (7.2) without external force. Nevertheless, there are not "physical" solutions for the interaction potential chosen. Thus we shall call them *virtual*.

$|\mathbf{f}^1\rangle$ verifies the following properties: (i) In a system with repulsive interaction, as we consider here, all the contact forces are compressive and therefore $f_{ij}^1 > 0$ for all contacts. (ii) It is well known from simulations and experiments that the distribution of contact forces is roughly exponential, or compressed exponential (see for example [50] for simulations in the frictionless case). This implies that the fluctuations of the contact forces are of order of the average value, leading to $\langle f_{ij}^1 \rangle^2 \sim \langle (f_{ij}^1)^2 \rangle = 1/N_c$ for a normalized force field. Thus we may introduce a constant c_0 such that:

$$\langle f_{ij}^1 \rangle = c_0 \frac{1}{\sqrt{N_c}} \tag{7.23}$$

Now we turn to the properties of the virtual forces $|\mathbf{f}^p\rangle$: (i) There are no physical constraints on the sign of the contacts forces for the virtual vectors. Furthermore, the $|\mathbf{f}^p\rangle$ must be orthogonal to $|\mathbf{f}^1\rangle$, whose signs of contact forces are strictly positive, and where the fluctuations in the contact compression is small. Therefore the virtual force fields have roughly as many compressive as tensile contacts.

7.4.2 Implementation of global strain

In our framework it turns out to be convenient to study the response to shear or compression as there are generally implemented in simulations. When periodic boundary conditions are used, an affine strain is first imposed on the system. Then the particles are let to relax. In general the affine strain is obtained by changing the boundary condition. Consider a 2-dimensional system with periodic boundary: it is a torus. For example a shear strain can be implemented by increasing one of the principal radii of the torus and decreasing the other. Then the distance between particles in contact

increases or decreases depending on the direction of the contact. In fact, this procedure of change of boundary conditions is formally equivalent to a change of the metric of the system. If the metric is changed from identity I to the constant metric $G = I + U$, the length of a vector $\delta\vec{l}_0$ becomes δl, such that $\delta l^2 = \delta\vec{l}_0 \cdot G \cdot \delta\vec{l}_0$. From this expression one deduces the change of distance between two particles is given by the formula:

$$\delta r_{ij} = \vec{n}_{ij} \cdot U \cdot \vec{R_{ij}} \tag{7.24}$$

Near jamming $\vec{R_{ij}} \approx \vec{n}_{ij}$ and therefore $\delta r_{ij} \approx \vec{n}_{ij} \cdot U \cdot \vec{n}_{ij}$. Formally, such a change of metric is strictly equivalent to a change of the rest length of the interactions with $y_{ij} = \vec{n}_{ij} \cdot U \cdot \vec{n}_{ij}$. Incidentally Eq.(7.16) can be used to compute the energy of such strain.

7.4.3 Compression

For a compression $U = -\epsilon I$ where I is the identity matrix and ϵ is the magnitude of the strain. Eq.(7.16) becomes:

$$\delta E = \frac{1}{2}(\sum_{ij} -\epsilon f_{ij}^1)^2 + \frac{1}{2}\sum_{p=2,..\frac{N\delta z}{2}} (-\epsilon \sum_{ij} f_{ij}^p)^2 \tag{7.25}$$

In the first sum all the terms have the same signs for a purely repulsive system, and this term leads to the strongest contribution. We have:

$$\delta E \geq (\sum_{ij} -\epsilon f_{ij}^1)^2 = \epsilon^2 (N_c \langle f_{ij} \rangle)^2 = \epsilon^2 c_0^2 N_c \tag{7.26}$$

On the other hand, δE is certainly smaller than an affine compression whose energy also goes as $\epsilon^2 N$. Therefore we find that:

$$\delta E \sim N\epsilon^2 \tag{7.27}$$

$$B \equiv \frac{\delta E}{N\epsilon^2} \sim \delta z^0 \tag{7.28}$$

The bulk modulus of an harmonic system jumps from 0 in the "gas" phase toward a constant when the system becomes jammed, as verified in the simulations. From this result follows that $p \sim (\phi - \phi_c)$. Note that this result holds only for purely repulsive systems. If there are as many tensile and compressive contacts, one recovers for the bulk modulus the result valid for the shear modulus, that we derive in the next section.

7.4.4 Shear

If a pure a shear strain is imposed, the tensor U is traceless. Let be ϵ the largest eigenvalue (in absolute value). The change of distance of two particles in contact due to shear $\delta r_{ij} = \vec{n}_{ij} \cdot U \cdot \vec{n}_{ij}$. It is a number of zero average if the system is isotropic, and fluctuates between $+\epsilon$ and $-\epsilon$ depending on the orientation of $\vec{R}_{ij}$. Eq.(7.16) becomes:

$$\delta E = \frac{1}{2} \sum_{p=1,..\frac{N\delta z}{2}} (\sum_{ij} f_{ij}^p \delta r_{ij})^2 \tag{7.29}$$

Each term in the summation gives on average:

$$\langle (\sum_{ij} f_{ij}^p \delta r_{ij})^2 \rangle = \sum_{ij} \langle (f_{ij}^p)^2 \delta r_{ij}^2 \rangle + \sum_{mn \neq ij} \langle f_{ij}^p f_{mn}^p \delta r_{ij} \delta r_{mn} \rangle \tag{7.30}$$

$$= \sum_{ij} \langle (f_{ij}^p)^2 \rangle \langle \delta r_{ij}^2 \rangle + \sum_{mn \neq ij} \langle f_{ij}^p f_{mn}^p \delta r_{ij} \delta r_{mn} \rangle \tag{7.31}$$

In principle they can be spatial correlations in the contact force amplitudes that leads to $\langle f_{ij}^p f_{mn}^p \delta r_{ij} \delta r_{mn} \rangle \neq 0$ even if $mn \neq ij$. Nevertheless we expect these terms to be negligible, as we argue at the end of the paragraph. Concerning the diagonal terms, one has $\delta r_{ij}^2 \approx \epsilon^2$ while $\sum (f_{ij}^p)^2 = 1$ by construction. Thus each term in the p summation is of order $\epsilon^2 \cdot 1$, and:

$$\delta E \sim N \delta z \epsilon^2 \tag{7.32}$$

which implies:

$$G \equiv \frac{\delta E}{N \epsilon^2} \sim \delta z \tag{7.33}$$

This is in agreement with the observation of [50].

We come back to the subtle question of the possible presence of spatial force correlations. In the simulation of soft spheres of [50], the spatial correlations of the real force field $|\mathbf{f}^1\rangle$ vanish for distance larger than roughly two particle diameters at the jamming transition. This property is restituted by simple models of force propagation such as the scalar q-model [84]. In this model, the system is decomposed in layers, and forces propagate downward from the top of the system to the bottom. Each particle as 2d contacts, with d particles of the upper layer and d particles of the layer below. The force field is builded recursively, following a local rule: when a particle receives from an upper layer a total force amplitude $f = f_1 + ... + f_d$, this force is distributed randomly to the d contacts below. This model mimics the fact that in an isostatic system external forces imposed at the top of a system propagate downward, as we discussed in Chapter 2. The randomness

of the force splitting sketches the randomness of the local configuration of contacts. This model presumably captures some of the relevant physics, as it leads both to a rather realistic force distribution, and does not display any spatial correlation in the force amplitude. Extending this model, we argue that when $\delta z > 0$ the typical virtual force fields do not display long-range correlations. Indeed following the same line of thought a typical virtual force fields can also be builded recursively with a similar local rule, the only difference being that some particles have more than 2d contacts. Thus the level of randomness increases as there are sometimes more ways to split the force between the different contacts below, accounting for the fact that there are many possible virtual force fields. Such an increase of randomness will certainly not create long-range correlation, and this simple model justifies our assumption that such correlations are negligible. Note that this model does not preclude that a few force fields can display long-range correlations, but simply supports that such correlations do not occur for the bulk part of the set of force fields. In particular, we expect that when $\delta z > 0$, the real force field $|\mathbf{f}^1\rangle$ displays long range correlations. Indeed this force field is solution of the whole elastic problem, and cannot be builded from any local rule. We expect that at large distances r, the correlation in force amplitude would follow $\langle f^1(0) f^2(r)\rangle \sim r^{-d}$ as in a continuous elastic medium with random disorder [83]. Accounting for this particular force field in Eq.(7.30) leads to a relative correction of G that goes as $\frac{\ln(N)}{N}$ which vanishes at the thermodynamic limit. Finally, note that some subset of these virtual force fields were studied in simulations [85], and that no long range correlations were noticed [86].

7.5 Discussion: non-affine displacements and length scales.

An affine shear costs an energy comparable to a compression. Thus the non-affine displacements that follow a pure shear diminish the cost in energy by a ratio that diverges at the jamming transition. Such non-affine displacements $|\delta\mathbf{R}_{\mathbf{n.a.}}\rangle$ are simply the displacements corresponding to the minimization of Eq.(7.11). It follows that $|\mathbf{y}^{\perp}\rangle = \mathcal{S}|\delta\mathbf{R}_{\mathbf{n.a.}}\rangle$. $\mathcal{S}$ can be made invertible if it is restricted to its image space, and we may write:

$$|\delta\mathbf{R}_{\mathbf{n.a.}}\rangle = |\mathcal{S}^{-\mathbf{1}}\mathbf{y}^{\perp}\rangle \tag{7.34}$$

A surprising recent observation was made in the Lennard-Jones simulations of [6, 7]: a rather large length scale (~ 30 particle sizes) appears in the correlations of the non-affine displacements. Eq.(7.34) is a linear equation. Any

strain $|\mathbf{y}\rangle$ can be decomposed as the sum of individual contact deformations. Therefore the non-linear displacements that follow a shear or a compression can be written as the sum of the responses to a contact deformation. The formalism of the present Chapter does not give any access to the spatial behavior of the response to a local deformation when $\delta z > 0$. On the other hand the results on the vibrational modes of Chapter 4 might bring interesting insights. Since any deformation can be decomposed on the vibrational modes, we expect the characteristic lengths l^* and l_t to characterize the non-affine displacements that follow a global strain. It would be useful to study these questions numerically near the jamming threshold.

8. Granular matter and Glasses

In the previous Chapters we discussed some properties of weakly-compressed harmonic soft spheres. Our main results are that (i) anomalous modes appear at low-frequency. They are related to the soft modes of floppy subsystems and are characterized by a length scale l^* (ii) there is a frequency scale ω^* below which the system does not behave as a continuous medium, but as in an isostatic state. In this Chapter we study the applicability of these results to more realistic systems. We first discuss granular matter, then glasses. The sections contained in this Chapter, apart the last one, can be read independently since they deal with distinct matters.

At least two modifications are necessary to apply our results to granular matter: (i) the presence of friction. Up to now we studied mostly radial interactions, and covalent bonds in Chapters 2 and 3. In section 8.1 we shall see that there is no conceptual difference when friction is present, apart from a change in the equations that define the soft modes. (ii) the potential used: In three dimensions grains do not interact harmonically. Modeling the contact with a Hertzian potential is more realistic, even thought it might not be perfect [87]. It corresponds to $\alpha = 5/2$ in Eq.(1.1). In section 8.2 we compute the density of states for such non-harmonic interactions. Our results agrees with the numerical findings of [50].

Then we discuss the vibrations of glasses. All glasses present attractive forces, such as Van der Waals interactions. Therefore each particle a priori interacts with all other ones, and the coordination number is not-well defined. In section 8.3 we show how one can deal with this problem and compute the density of states in simple cases where there is a strong hierarchy in the contact stiffness. Our results apply for systems as silica, where the covalent interactions inside the tetrahedron are the strongest as discussed in Chapter 4. In section 8.4 we consider systems without any clear-cut hierarchy in the distribution of contact stiffness, such as Lennard-Jones systems. The question is to know under which conditions our results on the vibrations of weakly-connected system can apply in such situations. We propose an improved variational argument which uses the distribution of the contact stiffnesses to evaluate the density of states. This leads to testable predictions on the nature of the excess-modes in such systems. In particular, this

should enable to decide whether or not the length scales that appears in the vibrations, the responses to a point force or the non-affine displacements in some Lennard-Jones simulations [6, 7, 8, 9] corresponds to the length scales we introduced in this thesis.

Let us specify that in this Chapter we shall not consider the effect of initial stress: this is a separated issue, which can be treated independently, as shown in Chapter 5.

8.1 Particles with friction

In this section we show how our results can be extended to particles with friction. We shall assume that all the contacts lie *inside* the Coulomb cone: particles in contact do not slide irreversibly for an infinitesimal perturbation. This means that there is no plastic deformation at the surface of contact: the contacts act as if they were "welded". Simulations showed that the validity of this assumption depends on the system preparation. The distribution of the contact tangential force was observed to vanish on the edges of the Coulomb cone in [53]. In [88], this was also observed if the dynamics that leads to jamming is fast. Otherwise, a finite fraction of the contacts was observed to lie on the Coulomb cone. For concreteness we consider the two dimensional case where the notations are simpler. The same ideas apply in higher dimensions. We consider elastic discs, whose repulsive interaction is harmonic: $\alpha = 2$. We add a term of friction, function of the shearing of the contacts. If two particles are adjacent, and one of them rotates an angle $\delta\theta$ while the other is rigidly pinned, the energy stored can be written in the form:

$$\delta E = \gamma \delta\theta^2 \tag{8.1}$$

γ describe the energy associated with the shearing of a contact. This can be easily generalized in 3-dimensions [89].

If a small displacement $|\delta\mathbf{R}\rangle$ and a small rotation field $\delta\theta = \{\delta\theta_i\}$ are imposed, the shear between two particles in contact is characterized by the following angle:

$$\delta\theta_{ij} = \theta_i + \theta_j + (\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij}^{\perp} \tag{8.2}$$

where $\vec{n}_{ij}^{\perp}$ is obtained by rotating $\vec{n}_{ij}$ by $+\frac{\pi}{2}$. Therefore, using (8.1), we can write for the expansion of the energy:

$$\delta E = \sum_{ij} \frac{1}{2}[(\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij}]^2 + \gamma[\theta_i + \theta_j + (\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij}^{\perp}]^2 + O(\delta\vec{R}^3) \tag{8.3}$$

For each contact ij there is a compression force f_{ij} and a tangential force $f_{ij}^{\perp}$. The system is at equilibrium when the force and the momentum are balanced on every particle:

$$\sum_j f_{ij}\vec{n}_{ij} + f_{ij}^{\perp}\vec{n}_{ij}^{\perp} = \vec{F}_i \quad \text{for all i} \tag{8.4}$$

$$\sum_j f_{ij}^{\perp} = \mathcal{M}_i \quad \text{for all i} \tag{8.5}$$

where $\mathcal{M}_i$ is the external momentum applied on particle i. This set of linear equation has $2N_c$ degrees of freedom and $3N-3$ constraints in two dimensions. Therefore a jammed system with friction that can sustain any generic external force field must be such that $2N_c > 3N-3$. At the isostatic point the coordination number is given by:

$$z_c = 2N_c/N \to 3 \tag{8.6}$$

In three dimensions one finds similarly $z_c = 4$. The soft modes are modes of null energy. In two dimensions Eq.(8.3) gives:

$$(\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij} = 0 \tag{8.7}$$

$$\theta_i + \theta_j + (\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij}^{\perp} = 0 \tag{8.8}$$

Again this is a linear system. Now there are $3N-3$ degrees of freedom, and $2N_c$ constraints. Following the procedure of Chapter 3 and 4 one can build anomalous modes that appear at a frequency $\omega^* \sim \delta z$, where $\delta z = z - z_c$, z_c being the isostatic limit with friction. This result was recently observed numerically in [90]. Furthermore, it is easy to show that this system (8.7) is, as in the frictionless case, the dual of (8.4). The results on the dipole of force and the scaling of the elastic moduli of Chapter 7 can then be recovered.

Note that contrarily to the frictionless case, this system does not need to be isostatic when the pressure vanishes. As we discussed in Chapter 2 and 3, in frictionless systems the two conditions that (i) the system must be rigid and (ii) the particles cannot interpenetrate lead to the unique solution for the coordination number: $z = z_c$. In the frictional case, these two antagonist requirements do not lead to a unique solution. The geometrical requirement that forbids the interpenetration of particles does not change, whereas the condition of rigidity becomes less demanding, as we just showed. In two dimensions one finds $3 \leq z \leq 4$, and in three dimensions $4 \leq z \leq 6$. Consequently the coordination of stiff frictional particles depends on the preparation of the system [88, 53]. In three dimensions when the friction coefficient $\mu \approx 1$ the simulations of [53] leads to $z \approx 4.5$ that implies $\delta z \approx 0.5$. We guess that this corresponds to a l^* of few tens of particle sizes, which could be probed by computing the vibrational modes of this system.

8.2 Extension to non-harmonic contacts

Up to now we considered harmonic interactions. Here we discuss the generalization of our argument to other contact potentials. Ref.[50] explored several other potentials, in particular the Hertzian contact potential describing the compressive energy of two elastic bodies. It corresponds to $\alpha = 5/2$ in Eq. (1.1). Ref.[50] observed a plateau in the density of states whose height D_0 scales as $p^{-1/6}$. They also observed a cutoff frequency ω^* varying as $p^{\frac{1}{2}}$. In the Hertzian case the quadratic energy of Eq. (2.1) becomes:

$$\delta E = \frac{1}{2}\sum_{\langle ij\rangle}(1-r_{ij})^{\frac{1}{2}}[(\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij}]^2 \tag{8.9}$$

The new factor $(1-r_{ij})^{\frac{1}{2}}$ amounts to a spring constant k_{ij} that depends on compression. The contact force $f_{ij} = \partial\delta E/\partial r_{ij}$ evidently varies as $(1-r_{ij})^{3/2}$. In what follows we start by neglecting the fluctuations that exist between the stiffnesses of the contacts. This treatment is sufficient to recover the scaling results of [50]. Then we discuss how such fluctuations can be taken into account to improve the bound on the density of states derived in Chapter 3.

The new factor $V''(r_{ij}) = (1-r_{ij})^{\frac{1}{2}}$ rescales the energy. To account for this overall effect, we replace $(1-r_{ij})^{\frac{1}{2}}$ by its average $\langle(1-r_{ij})^{\frac{1}{2}}\rangle$. Expressed in terms of contact forces, this factor is proportional to $\langle f_{ij}^{1/3}\rangle$. Again replacing f_{ij} by its average, the factor becomes $\langle f_{ij}\rangle^{1/3}$. This average is related to the pressure p, via $p \approx \langle f_{ij}\rangle$. Thus in this approximation the overall effect is to rescale the energy by a factor $k(p) \sim p^{1/3}$.

$$\delta E = \frac{k(p)}{2}\sum_{\langle ij\rangle}[(\delta\vec{R}_j - \delta\vec{R}_i).\vec{n}_{ij}]^2 \tag{8.10}$$

Apart from this prefactor, the energy and the dynamical matrix are identical to the harmonic case treated above. Thus each normal mode frequency gains a factor $k^{\frac{1}{2}} \sim p^{1/6}$. In the harmonic case the crossover frequency follows $\omega^* \sim \delta z$. In the Hertzian case, it too gains a factor $k^{\frac{1}{2}}$, so that $\omega^* \sim k^{\frac{1}{2}}\delta z$. When the initial stress is taken into account, the bound on the lowest-frequency anomalous modes ω_{AM} still has the form

$${\omega_{AM}}^2 \leq \omega^{*2} - A_2 p \tag{8.11}$$

For a marginally-stable system we still have $\omega_{AM} \approx 0$ which leads to an unaltered relationship between ω^* and p, namely $\omega^* \sim p^{\frac{1}{2}}$. Comparing with our previous estimate of ω^* yields $\delta z \sim p^{1/3}$. Furthermore, the plateau density of states D_0 has the dimensions an inverse frequency, and thus gains

a factor $p^{-1/6}$. Since the harmonic D_0 had no dependence on p, the Hertzian $D_0(p)$ also should vary as $p^{-1/6}$. The scaling behaviors seen in [50] agree with these expectations. These arguments may be applied to general values of the interaction exponent α.

Additional effects could in principle alter the low-frequency modes in the Hertzian case. When harmonic springs are replaced by Hertzian springs, the contacts supporting different forces have different stiffnesses. The variational argument derived in Chapters 3 and 4 does not consider such fluctuations. If it is applied as is to the more general case with fluctuations, the energy of the corresponding anomalous modes defined in Eq.(3.1) simply gains a factor $\langle k \rangle$, the average stiffness. This can be deduced from Eq.(3.3) by neglecting the correlations between the soft modes displacements and the stiffnesses amplitudes[19]. Thus, this variational argument corresponds to the simple derivation of the previous paragraph where quantities are replaced by their average. Here we propose to use the stiffness fluctuations to improve this variational argument by a numerical factor. Instead of modulating the soft modes by $\sin(\frac{x_i \pi}{L})$ to obtain the anomalous modes of Eq.(3.1), we introduce a more general phase $\psi(i)$ and modulate the soft modes with $\sin(\psi(i))$. Then we minimize the average energy of these anomalous modes with respect to ψ, imposing $\psi = 0$ and $\psi = \pi$ on the boundaries $x = 0$ and $x = L$. The expression of the energy corresponds to Eq.(3.3) with $\sin(\psi(i))$ replacing $\sin(\frac{x_i \pi}{L})$ and each term of the sum multiplied by a stiffness k_{ij}. When averaged on the soft modes amplitude, one obtains the average energy of the anomalous modes $\langle \delta E \rangle \sim \sum_{ij} k_{ij} \cos^2(\psi(i))[\psi(i) - \psi(j)]^2$. This has to be minimized with respect to the variables $\psi(i)$. We expect the phase ψ to vary slowly in space, and not to differ dramatically from our previous solution $\phi = \pi x/L$. Hence for a large system we may consider that the cosines terms that appear in the energy sum are constant locally. Then the local minimization of the energy corresponds to the problem of conductivity in a random network of resistors where $\delta E = \sum_{ij} {r_{ij}}^{-1}[U(i) - U(j)]^2$, where r_{ij} is the resistor between the vertex i and j, and $U(i)$ is the Coulomb potential in i. Therefore k_{ij} corresponds to the inverse of a resistance r_{ij}, and $\psi(i)$ to the potential $U(i)$. Effective medium theory [91] furnishes a good approximation of the conductivity, and therefore of the energy δE, of such a random system. For example if the network of contacts is sketched by a square or cubic lattice, one obtains for the effective stiffness the following equation:

$$\langle \frac{k_{eff} - k}{(d-1)k_{eff} + k_{ij}} \rangle = 0 \tag{8.12}$$

[19] We expect such correlations to be very small, as the amplitudes of the stiffnesses do not enter in the soft modes equation.

where the average is taken on the distribution of stiffness. Because of the convexity of the inverse function, one obtains that $k_{eff} \leq \langle k \rangle$. Expectedly, the energy in this improved variational argument is smaller than in our previous result, since it is proportional to k_{eff} instead of $\langle k \rangle$. For the distributions of stiffnesses we are considering here [20] where we expect no delta function in $k = 0$, nor any fat tail at large k, k_{eff} as computed in Eq.(8.12) and $\langle k \rangle$ are of the same order of magnitude. Thus the bound on the density of states is improved by a numerical factor when this improved argument is used.

8.3 $D(\omega)$ in systems with various interaction types

In this section we study the density of states of systems where several types of interactions are at play, and where the amplitude of the strongest interaction is much larger than the others. In particular we think about tetrahedral covalent networks such as silica. The strong interaction corresponds to the deformation of the tetrahedra, which is much larger than the bending of the Si-o-Si bonds or the Van der Waals attractions, as we discussed in Chapter 3. In Chapter 3 and 4 we showed how to compute the density of states of weakly-connected networks with only one type of interaction: no weak interactions were present. The density of states is then described by a plateau that appears above a frequency $\omega^* \sim \delta z$. In this section we show how to treat the weak interactions by simple perturbation. We simply consider the anomalous modes that would appear at ω^* if the weak interactions were not there. Then we compute the change in energy induced by the weak interactions on these anomalous modes. If these modes are on average shifted by an energy ζ, the plateau in the density of states appears at a frequency $\sqrt{\omega^{*2} + \zeta}$. In what follows we show how to evaluate ζ and discuss some examples.

8.3.1 Model

Consider for concreteness a system with radial interactions, where "strong" interactions with stiffness close to unity form a rigid network of coordination $z > z_c$, and where the weak interactions have stiffnesses of order $\eta \ll 1$.

[20] As we discussed in Chapter 7, the potential does not enter in the computation of the force field at the isostatic threshold. Thus if the configurations obtained at threshold are similar for the different potentials α, as it seems to be the case, the same force distribution is obtained for any α. Such force distribution is known to be roughly exponential. From this we can deduce the distribution of stiffnesses $P(k) \sim k^{\frac{1}{\alpha-2}} e^{-k^{\frac{\alpha-1}{\alpha-2}}}$.

Consider the lowest-frequency anomalous modes that appear at ω^* in the rigid network of coordination z. The energy correction ζ induced by the weak interactions on such mode is:

$$\zeta = \frac{1}{2}\sum_{\langle ij\rangle} \eta_{ij}[(\delta\vec{R}_i - \delta\vec{R}_j)\cdot\vec{n}_{ij}]^2 \tag{8.13}$$

where the sum is taken on all pairs of particles interacting with the weak interaction of stiffness η_{ij}. Here we consider that all pairs of particles $< ij >$ that appears in this sum are not in contact through the strong interaction, since such terms would simply re-normalize the strong stiffnesses, and can be taken into account when ω^* is computed in the first place. To estimate $\langle\zeta\rangle$ we need to compute $\langle[(\delta\vec{R}_i - \delta\vec{R}_j)\cdot\vec{n}_{ij}]^2\rangle$. If the displacement of particles i and j were uncorrelated, one would get for a normalized mode $\langle[(\delta\vec{R}_i - \delta\vec{R}_j)\cdot\vec{n}_{ij}]^2\rangle = \frac{2}{dN}$. We expect this to be the case when the distance r_{ij} between the two particles is large. When the distance between the pair ij diminishes, the displacement of particles i and j becomes correlated. Thus we may write $\langle[(\delta\vec{R}_i - \delta\vec{R}_j)\cdot\vec{n}_{ij}]^2\rangle = \frac{2c(r_{ij})}{dN}$, where $c(r_{ij}) < 1$ characterizes the correlations at distance r_{ij}. Since we are only considering pairs of particles which are not in contact for the rigid network, we expect that for the anomalous modes of such a network considered here, the correlation of the displacement of particles i and j is "weak", that is $c(r_{ij})$ is bounded below by a constant c_0 of order one. This is supported by our numerical result of Fig.(5.1), that shows that if two displacement degrees of freedom are not fixed by the soft mode condition, then the correlation between these displacements is weak, even when these two particles are close. At the level of approximation we are considering here we shall simply assume that $c(r_{ij}) = 1$. Then we obtain:

$$\zeta = \sum_{ij} \eta_{ij}\frac{1}{dN} \equiv \frac{1}{d}\int \rho(\eta)d\eta \tag{8.14}$$

where we introduced the number $\rho(\eta)$ of pairs with stiffness η per particle per unit stiffness.

8.3.2 Density of states of square and cubic lattices

In what follows we discuss in particular the situation where the rigid network is isostatic $z = z_c$. Our present argument shows that a plateau appears in the density of states at $\omega \sim \sqrt{\zeta}$. To illustrate this idea we consider the example of the square (or cubic) lattice of point particles where first neighbors interact with spring at rest, see Fig.(8.1). Our argument of Chapter 3, which does not depend on disorder but only on coordination, applies to this system.

The square and the cubic lattice are isostatic, therefore there density of states is constant. In this particular case the soft modes are trivial one-dimensional object: they are the independent translations of the columns. Thus for these systems a simpler way to prove that the density of states is constant, instead of going through the argument of Chapter 3, is to consider such lattices as an assembly of disconnected one dimensional lattices, which obviously have a constant density of states. One may ask what happens to the density of states of such system if a weak interaction is added to couple these independent lines. As an example, we consider that the second neighbors are now coupled: we add springs with small stiffness η as drawn in Fig.(8.1). ρ is then a delta function at stiffness η, of amplitude 2 if one considers the square lattice (there are two weak interactions per particles). Therefore our rough evaluation for the appearance of the plateau of Eq.(8.14) yields $\omega = \sqrt{\eta}$. Since this system is crystalline, the normal modes are plane waves and the density of states can be computed exactly. The density of states display a cross over between a Debye regime and a plateau. A direct estimation of the cross-over energy gives $\delta E \approx 2\eta$ which leads to $\omega = \sqrt{2\eta}$ [21]. A cross-over energy is arbitrarily defined, here we simply note that our two estimates are similar. Finally, note that our argument indicates that the cubic lattice are very floppy if the second neighbor interactions are small in comparison with those of the first neighbors. This happens if the interactions potential decays rapidly. In practice, there are no simple elements that crystallize in a simple cubic crystal [25] because such structures are too floppy. For example it is not possible to crystallize a 12-6 Lennard-Jonnes in a square lattice even at zero pressure and zero temperature, because such a structure is mechanically unstable [22]. In general simple cubic crystals have charged particles, as in NaCl, and the second neighbors interactions are not negligible.

[21] The frequency of a plane wave of wave vector $\vec{\kappa}$ can be written $\omega \sim a(\vec{\kappa}/\kappa)\sin(\kappa)$, where the function $a(\vec{\kappa}/\kappa)$ only depends on the direction of the wave vector. If the sine is approximated by an affine function on obtains $\omega \sim a(\vec{\kappa}/\kappa)\kappa$. Then it is straightforward to show, by summing on all directions $\vec{\kappa}/\kappa$, that in this approximation the density of states is linear (as expected for a 2-dimensional crystal) until a frequency ω_0, defined as the lowest frequency at which some plane waves reach the edge of the Brillouin zone. These plane waves have the smallest $a(\vec{\kappa}/\kappa)$. In this example they correspond to the transverse waves whose wave vector is in the direction of the axis x or y as indicated in Fig.(8.1). When the wave vector reach the boundary of the Brillouin zone, the energy of such waves can be computed. One finds $E_0 = 2\eta$.

[22] Imposing that the pressure is zero brings a condition between the forces carried by the first neighbors and by the second neighbors. In a 12-6 Lennard-Jonnes it can be satisfied only if the second neighbors are located behind the inflection point of the potential. Therefore the second neighbors have a negative stiffness, which destabilizes the marginally stable network constituted by the first neighbors.

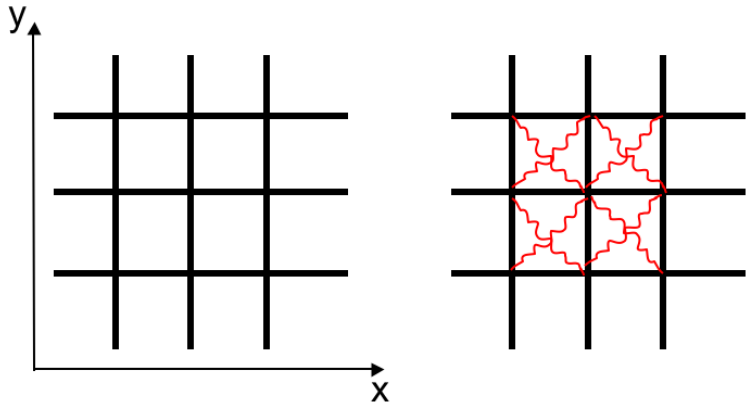

Fig. 8.1: Left: Square lattice with interaction between first neighbor of stiffness 1. Right: the same lattice where interactions of stiffness $\eta \ll 1$ have been added between second neighbors.

8.3.3 The boson peak of silica

Silica is the most common glass, much studied in experiments and in simulations. It is also known to have one of the strongest excess of low-frequency modes, or boson peak, see [46] for a review of empirical results and models. In this paragraph we propose an explanation for its density of states at low-frequency, and for the nature of the excess-modes. As we discussed in Chapter 4, the strongest interactions in silica lies inside the tetrahedron SiO_4. If the other interactions are neglected, one can model the glass as an assembly of linked tetrahedra of appropriate stiffness loosely connected at corners, the "rigid unit modes" model [70]. This description is supported by recent Hyper-Raman scattering experimental results that show that the low-frequency excess modes correspond to the motion and rotation of stiff tetrahedra [93]. As we discussed in Chapter 4 such a tetrahedral network has a constant density of states at low frequency. Following the results of the last section, one may evaluate the effect of the weaker interactions (in particular the bending of the Si-o-Si bond and the Van der Waals interaction) on the density of states by computing the energy shift ζ that they induce on the anomalous modes. ζ can in principle be estimated from the knowledge of the distribution of the stiffnesses of the weak interactions ρ. Using the stiffness of the Si-O-Si bending interaction obtained *ab initio* [72], and the molecular mass, we obtain a frequency 1.4 Thz, which estimates the order of magnitude of $\sqrt{\zeta}$. According to the previous arguments of the present section, we expect thus silica glass to display a plateau in its density of states that should

appear at a frequency of the order of 1 Thz. This is indeed what is observed in simulations: silica glass present a well-defined plateau in the density of states, which appears at a frequency corresponding to the boson peak, see e.g. Fig.15 of [94] or Fig.(8.2) for numerical results.

Our argument does not involve disorder. Thus it must also apply for the crystals of the same composition and similar densities such as the α and β-cristobalite, since these crystalline structures are formed, as silica, by SiO_4 tetrahedra connected at the corners. β-cristobalite has the structure of the diamond, in which the tetrahedra correspond to the 4 carbons bonded to a central carbon, whereas α-cristobalite has a tetragonal structure. Empirically a boson peak is observed in all these materials [95]. Numerically, a plateau indeed appears in $D(\omega)$ at roughly the same frequency in the cristobalite α and β [68] and in the glass, as shown in Fig.(8.2).

More generally, there are crystals showing an excess density of states at frequencies that correspond to the typical boson peak frequency in glasses [46, 47, 48, 49]. This implies that the disorder is not a necessary condition to obtain excess modes. This is a crucial point as in many theories of the boson peak disorder is the only relevant parameter. Such theory certainly cannot explain the density of states of silica, very similar to the cristobalites. Rather, we argue here that the key feature that determines the density of states is weak coordination.

Note that if disorder is not relevant to compute the density of states, it affects the nature of the vibrational modes. For example properties such as transport are very different between silica glass, and the cristobalites. Thus the peculiarity of the amorphous state lies in the *nature* of the excess-modes, not in the density of states [46]. It is useful to note the parallel between cristobalite and silica glass on the one hand and cubic lattice and the jamming threshold of elastic spheres on the other hand. In both cases the amorphous solid and the crystal have a similar density of states, but the anomalous modes in the amorphous phase are not plane waves. Disorder strongly affects the anomalous modes, and makes them very heterogeneous, as it appears in Fig3.1.

There is an interesting geometrical parallel between a tetrahedral network and an assembly of spherical particles with friction. The equation that defines the soft modes of these two systems are the same, and corresponds to the extension of Eq.(8.7) in three dimensions. This comes from the fact that when friction is present, the contacts can be considered as "welded", as we discussed in 8.1. Similarly in a tetrahedral network the corners of each tetrahedra in contact are attached. In both cases the soft modes corresponds to the rigid motion of the particles that keep the contact point welded.

Finally, we considered systems in which the strong interactions form a

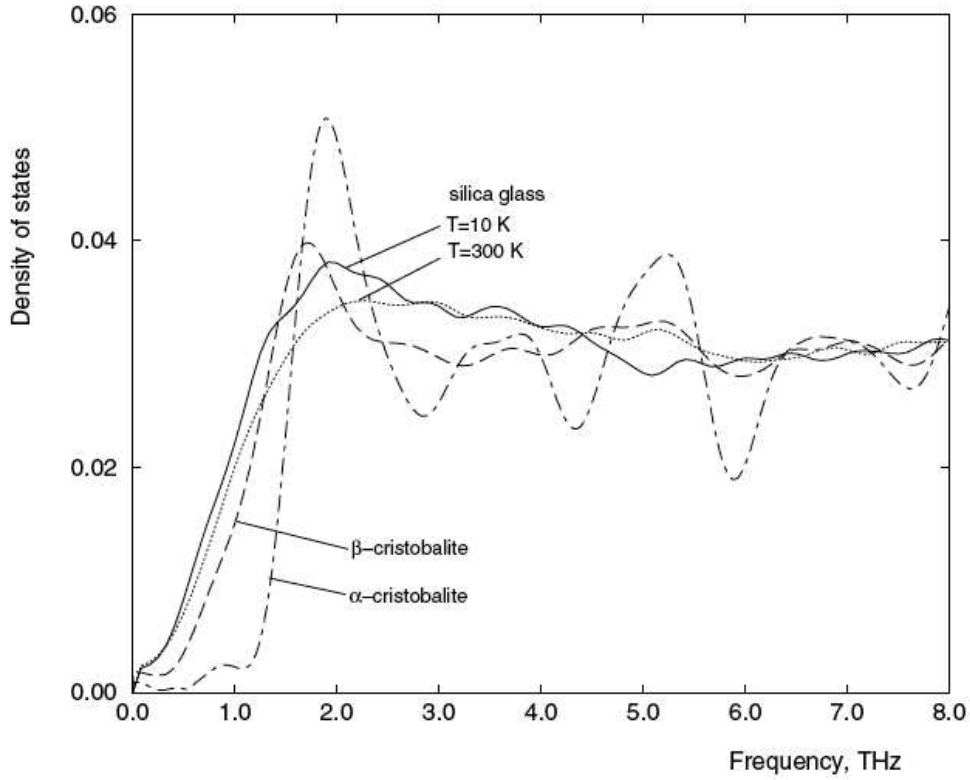

Fig. 8.2: Density of states of silica glass (at temperatures of 10 and 300 Kelvins), α-cristobalite and β-cristobalite. This figure is taken from the simulations of [68].

rigid network, that is $\delta z > 0$. Our finding is that the density of states can be understood in terms of the anomalous modes that we introduced in Eq.(3.1). Note that if the concept of such anomalous mode is new, soft modes have been used in several different fields as we discussed in Chapter 2. When the network of strong interactions is not rigid, $\delta z < 0$, and the stiff network has got some soft modes. Consequently there is delta function in the density of states at $\omega = 0$ if the weak interactions are neglected. When the weak interactions are taken into account, Thorpe and collaborators [97] argued that these soft modes shift to higher frequency (in our notation by an amount $\sqrt{\zeta}$), leading to a peak in the density of states. The existence of such peak was observed in neutron scattering data of Ge_xSe_{1-x} glasses. When the composition of this glass is changed as x decreases, the covalent network becomes less and less connected. At $x = 0.24$ it is isostatic. The peak amplitude was observed to diminish as x decreases toward 0.24 [96, 97], as predicted.

8.4 Lennard-Jones systems

In this section we discuss the vibrations of systems where (i) coordination is not well defined a priori and (ii) there is not a clear-cut hierarchy in the amplitudes of interactions. We have in mind Lennard-Jones systems, or purely repulsive system with a potential of the form $V(r) = r^{-\beta}$. In particular, we discuss under which conditions such systems should display the anomalous modes we introduced in Chapters 3. One important motivation is the recent results on Lennard-Jones systems [6, 7, 8, 9] where excess-modes are observed, and where a characteristic length scale appears, in particular in the vibrations of the system and in the response to a local force. It is tempting to associate the excess-modes of [6] to the anomalous modes, and the length scale observed by the authors to the lengths we introduced in the previous Chapters. In this section we propose an improved variational argument, in the spirit of the last section, to evaluate the density of states of such systems. We shall argue that this variational method is efficient if the interaction potential decays rapidly. In this case anomalous modes appear characterized by a length scale. We propose a numerical test to check whether or not the excess-modes found in the Lennard-Jones system of [6] correspond to the anomalous modes we discussed.

For concreteness we consider a system with radial interactions, and introduce the density of stiffness $P(k)$, which denotes the number of pairs of particles whose interaction has a stiffness k per particle per unit stiffness. In the harmonic soft sphere system that we considered in the previous Chap-

ters $P(k) = \frac{z}{2}\delta(k=1)$. In the previous section, $P(k)$ could be decomposed in two well-separated distributions, one around $k=1$ of amplitude $z/2$ for the strong interactions, and one at much weaker k for the weak ones. Now we consider a general distribution $P(k)$. We propose the following family of variational arguments to define an effective coordination and evaluate the density of states. Let us introduce an artificial cut off $\lambda > 0$, such that the coordination number $z(\lambda) = z_c + \lambda$. We use this cut-off to decompose the interactions in two groups: the $z(\lambda)/2$ strongest ones, which form a network of " strong" interactions of coordination $z(\lambda)$, and the rest. This allows us to repeat the argument of the previous section: we first compute the energy $\omega^{*2}(\lambda)$ of the low-frequency anomalous modes of the "strong" network. Then we evaluate the correction in energy $\zeta(\lambda)$ induced by the rest of the interaction on these anomalous modes, as computed in the last section. Finally, one obtains the energy $E(\lambda) = \omega^{*2}(\lambda) + \zeta(\lambda)$ which gives a bound of the energy of appearance of the excess-modes. Then the standard procedure consists in minimizing $E(\lambda)$ with respect to λ. If λ_0 minimizes this quantity, then $\sqrt{E(\lambda_0)}$ gives the best estimate of the frequency of the anomalous modes. Furthermore, one can define an effective coordination $z = z_c + \lambda_0$. The length that characterizes the anomalous modes follows, according to Chapter 4, $l^* \sim \lambda_0^{-1}$. In what follows we compute $E(\lambda)$ and λ_0, discuss the quality of the estimate obtained by this argument, and compute other quantities that should enable to test if the excess-modes of Lennard-Jones systems are correctly described in terms of anomalous modes.

We first evaluate $\omega^{*2}(\lambda)$. In the harmonic case where all the stiffnesses are 1, we had $\omega^{*2} = A_1 \delta z$, where A_1 can be deduced numerically from the data of [50]. From Fig.4.1 one has in three dimensions $A_1 \approx 0.12$. In the network of coordination $z(\lambda)$, the amplitude of the stiffnesses fluctuates. To present this argument in its simplest form, we neglect these fluctuations, and replace the stiffness of each contact by the average stiffness as we discussed in the first part of section 8.2. We also showed there how a more accurate argument could be made, and we shall come back to this issue later. From our approximation follows that $\delta E \sim \langle k \rangle$, and in our notation we obtain $\omega^{*2}(\lambda) = A_1^2 \langle k \rangle_\lambda \lambda^2$, where $\langle k \rangle_\lambda = \frac{2}{z(\lambda)} \int_{k(\lambda)}^{\infty} k P(k) dk$. $k(\lambda)$ is the weakest stiffness of the network of "strong" interactions, defined as:

$$\int_{k(\lambda)}^{\infty} P(k) dk = \frac{z(\lambda)}{2} \tag{8.15}$$

.

We now compute $E(\lambda)$ and λ_0. From Eq.(8.14) we get for the correction

in energy ζ: $\zeta(\lambda) = \frac{1}{d}\int_0^{k(\lambda)} P(k)kdk$, and finally:

$$E(\lambda) = A_1^2\langle k\rangle_\lambda \lambda^2 + \frac{1}{d}\int_0^{k(\lambda)} P(k)kdk \tag{8.16}$$

The effective extra-coordination λ_0 is defined as $\frac{dE(\lambda)}{d\lambda}|_{\lambda=\lambda_0} = 0$. This equation defines a non-trivial minimum in general, since (i) $\frac{dE(\lambda)}{d\lambda}|_{\lambda=0} = -k(0) < 0$ as required by mechanical stability, and (ii) if the potential decays reasonably fast with distance, $E(\lambda) \sim \langle k\rangle_\lambda \lambda^2 \sim \lambda \to \infty$ as $\lambda \to \infty$. Using Eq.(8.15), the minimization of $E(\lambda)$ leads to:

$$2\lambda A_1^2\langle k\rangle_{\lambda_0} + A_1^2\lambda_0^2\frac{d\langle k\rangle_\lambda}{d\lambda} = \frac{1}{2d}k(\lambda_0) \tag{8.17}$$

A necessary condition for this variational argument to be relevant is that the solution λ_0 of Eq.(8.17) is small (at least smaller than the end of validity of the scaling of Fig.4.1, say $\lambda_0 < 2.5$ in three dimensions). Assuming that it is the case (and checking it self-consistently later), one may neglect the term quadratic in λ^2 in Eq.(8.17) to find:

$$\lambda_0 = \frac{1}{4dA_1^2}\frac{k(\lambda_0)}{\langle k\rangle_{\lambda_0}} \tag{8.18}$$

In three dimension using the value of A_1 one gets $\lambda_0 \approx 5\frac{k(\lambda_0)}{\langle k\rangle_{\lambda_0}}$. Thus a first requirement for this variational argument to apply is that $k(\lambda_0)$ must be several times smaller than $\langle k\rangle_{\lambda_0}$. A second requirement is that the bound of the energy $E(\lambda_0)$ must be relatively small in comparison with the Debye energy $E_D = \omega_D^2$: if this bound is too large, our variational argument does not lead to a correct estimate of the density of states, and the vibrational modes are not well described by anomalous modes. This implies that the correction in energy $\zeta(\lambda)$ must not be too large. It is clear that these two conditions are satisfied when the interaction potential decays fast enough. Consider for example a potential of the form $V(r) = r^{-\beta}$. If the potential decays slowly, $\beta > 0$ is small, then the two requirements we just evoked are not satisfied. In three dimensions for example there are on average 12 neighbors surrounding a particle. If β is small they all interact with a similar stiffness with the particle lying at the center, and one finds that $k(\lambda_0)$ is close to $\langle k\rangle_{\lambda_0}$, and not several times smaller. Furthermore the interactions with the second, third, etc... neighbors are not negligible either, and ζ is large. Thus our variational argument is not relevant in this case, and we expect the vibrations of such system to correspond to plane waves, and not to display excess-modes. If β is large enough, then the two requirements are satisfied if

the system is amorphous. To show that we assume that $g(r)$ does not evolve much with the potential considered, as it is generally observed with radial interactions in the glass phase. Then, when β becomes large, the hierarchy among the stiffnesses diverges: if one particles has 2 neighbors at distance r_1 and r_2 with $r_1 < r_2$, the ratio of the stiffnesses of these two interactions is $\frac{k_1}{k_2} = (\frac{r_2}{r_1})^{\beta+2}$ which diverges when β increases. It follows that ζ becomes negligible, and that $\lambda_0 \sim \frac{k(\lambda_0)}{\langle k \rangle_{\lambda_0}} \to 0$ as $\beta \to \infty$.

To decide whether or not the anomalous modes that appear in this variational argument are responsible for the excess-modes observed in Lennard-Jones systems, such as those of [6], one may consider a more precise observable that characterizes the anomalous modes. The quadratic energy of a mode is the sum the energy condensed in every contact $\delta E(k_{ij}) = \frac{1}{2} k_{ij} [(\delta \vec{R}_i - \delta \vec{R}_j) \cdot \vec{n}_{ij}]^2$. Consider the average energy condensed in a contact of stiffness k, that we denote $\langle \delta E(k) \rangle$, where the average is taken on all the contacts of similar stiffness k. According to our variational argument, for the lowest-frequency anomalous modes $\langle \delta E(k) \rangle$ varies as follow: if $k < k(\lambda_0)$, the displacement of the particles in contact are not correlated and $\langle \delta E(k) \rangle = \frac{1}{d} k$. If $k > k(\lambda_0)$, the interaction belongs to the rigid network. Then the relative longitudinal displacement between particles in contact corresponds to the modulation by a sine on a length scale $l^* \sim \lambda_0^{-1}$, therefore $\langle \delta E(k) \rangle \sim k \lambda_0^2$. Thus the curve $\frac{\langle \delta E(k) \rangle}{k}$ is a step function, which jumps at $k = k(\lambda_o)$, as represented in Fig.(8.3). In the next paragraph we shall argue that this result is not exact and that the step is not sharp, but smooth. In any case, if the excess-modes observed in Lennard-Jones systems are related to the anomalous modes describe here, they should exhibit this cross-over. This could be verified numerically. Furthermore, from this curve $k(\lambda_0)$ can be characterized: it is the stiffness at which the cross-over takes place. From $k(\lambda_0)$ one can compute the effective coordination λ_0 if the distribution of stiffness $P(k)$ is known.

To conclude, we expect that the present variational argument could be improved in several ways. For example the fluctuations of stiffnesses could be taken into account to estimate ω^*, as discussed at the end of section 8.2. This leads to corrections on the quantity $\langle \delta E(k) \rangle / k$ for $k > k(\lambda_0)$ that can be estimated using effective medium theory. One finds for the anomalous modes that the quantity $\langle \delta E(k) \rangle / k$ is not constant for $k > k(\lambda_0)$, but decreases with k, as represented in the full line curve of Fig.(8.3). Such effects smooth the step function drawn in Fig.(8.3).

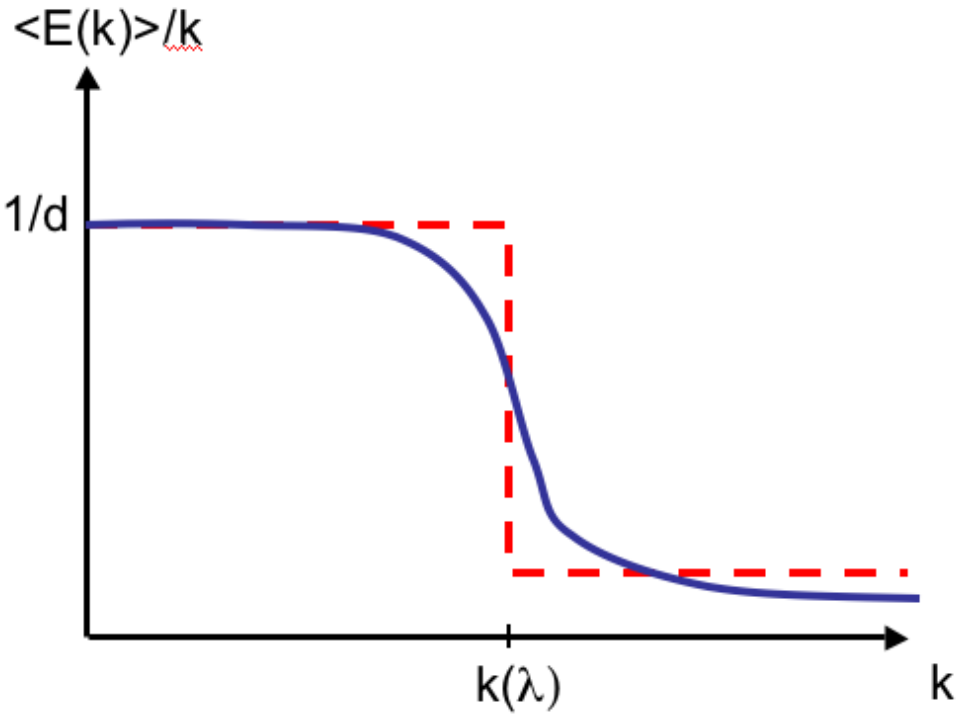

Fig. 8.3: $\langle \delta E(k) \rangle / k$ as defined in the text *vs* the stiffness k for an anomalous mode. The red dashed curve corresponds to the variational argument presented in the text. The black continuous curve is what we expect qualitatively when the variational argument is improved.

9. Rigidity of hard sphere liquids near the jamming threshold

There is something mystifying about the properties of a hard sphere system. On the one hand it is a simple system: no energy is involved, the only rule is that particles cannot interpenetrate, and only entropy matters. On the other hand, it displays a rich phenomenology. When the packing fraction ϕ is slowly increased from the liquid phase, it crystallizes. If the ϕ is increased rapidly, crystallization is avoided, and a glass transition is observed. The time τ_α that characterizes the de-correlations of the density fluctuations at some vector q grows rapidly. For a 3-dimensional monodisperse system τ_α becomes inaccessible numerically above $\phi_0 \approx 0.59$. Nevertheless the packing fraction can be increased further until the pressure diverges, when the distance between particles vanish: it takes place at $\phi_c \approx 0.64$, the random close packing. At packing fractions between ϕ_0 and ϕ_c the structure of the system is frozen, apart from the fast rattling of the particles around their average position.

The mode coupling theory furnishes predictions for the relaxation of the density fluctuations near the glass transition in rather good agreement with empirical data [98]. Nevertheless there is no clear understanding of the spatial nature of the events that relax the system, neither of the heterogeneous dynamics that has been observed near ϕ_0. A necessary first step to study such questions is to understand the cause of the rigidity of a hard sphere system at times scales $t \ll \tau_\alpha$. In the conventional picture [99], the freezing of the liquid at times $t \ll \tau_\alpha$ is interpreted with the "cage" effect. As the density increases, the cages formed by the neighboring particles tighten, and the characteristic time for a particle to escape its cage increases. This description considers the stability toward the motion of one single particle. It is dangerous since, as we discussed in the previous Chapters, the stability against *collective* motions of particles is more demanding than the stability against individual particles displacements. For example in d dimensions $d+1$ particles are enough to pin one particle. Nevertheless, a system with a coordination number $d+1$ is unstable to cooperative motion, as shown by the Maxwell criterion.

In this Chapter we study the microscopic cause of the glass rigidity. In

particular we study the hard sphere glass at packing fraction ϕ near ϕ_c, and we derive some elastic properties of the glass phase. Following the ideas of the previous chapters, we show that the solid-like behavior at $t \ll \tau_\alpha$ requires the formation of a rigid structure with a sufficiently large coordination number. We argue that such structure corresponds to the network that carries momentum between the particles, which was introduced recently to study granular flows [100]. Our main achievement is to compute an effective potential between particles in contact through this network. We show theoretically, after averaging over the fast fluctuations or rattling of the particles, that the hard sphere potential becomes logarithmic. This result is exact at ϕ_c, and is in very good agreement with a direct numerical check. This allows us to define normal modes and to apply the results we found for soft sphere solids near the jamming threshold. In particular, the extended Maxwell criterion applies, which yields an inequality for the network coordination. We confirm this result numerically. We compute the scaling of the high-frequency elastic moduli near ϕ_c. More generally, this approach shows that the jamming threshold acts as a critical point both in the solid and in the glassy liquid phase. This suggests original relaxation processes in the super-cooled phase.

9.1 Coordination number and force

For concreteness we consider a hard sphere system without dissipation where particles collide elastically. We show in the next section how to generalize our finding to the dissipative case where particles follow Brownian motion. We consider packing fractions ϕ such that the typical collision time between two neighbors τ_c is much smaller than τ_α. We introduce an arbitrary time t_1, much larger than the collision time, and much shorter than any time scales at which the structural relaxation occurs. Two particles are said to be in contact at a time t if they collide at least once in the time interval $[t - t_1/2, t + t_1/2]$. This enables us to define the coordination number z as $z = 2N_c/N$, where N_c is the total number of contacts. Then, we define the contact force $\vec{f}_{ij}$ as the total momentum per unit time exchanged between the two particles:

$$\vec{f}_{ij} = \frac{1}{t_1} \sum_{n=1}^{n=n_{col}} \Delta \vec{P}_n \tag{9.1}$$

where the sum is made on the total number of collisions n_{col} between i and j that took place in the time interval $[t-t_1/2, t+t_1/2]$, and $\Delta\vec{P}_n$ is the momentum exchanged at the nth shock. These definitions were first introduced in

a work on dense granular flows [100], and were used recently on hard sphere systems [79, 101]. Coordination and contact forces depend a priori on an arbitrary parameter t_1. In the high-packing fraction system we studied, we did not observe any relevant dependence of these objects with t_1 as long as $\tau_c \ll t_1 \ll \tau_\alpha$ [22].

In Fig(9.1) we show a two-dimensional example of the contact force field obtained with such procedure at packing fraction ϕ very close to ϕ_c. Note that the forces are roughly balanced on every particle, as it must be the case on time scales over which the structure is stable.

To obtain high packing fraction configurations such as the one of Fig.(9.1), we proceed as follows: we consider the two-dimensional polydisperse [23] configurations of [50] at the jamming threshold at packing fraction $\phi_c \approx 0.83$. At this packing fraction the particles are in contact. Then we reduce all the particles diameter of a relative amount ϵ. This leads to configuration of packing fraction $\phi = \phi_c(1-\epsilon)^2$. We assign a random velocity to every particle. A Newtownian dynamics is then computed using an event-driven simulation. As we shall discuss later, such a protocol does not lead to a system at thermal equilibrium. Nevertheless, we are not interested in thermodynamic properties, and in practice systems with such high packing fraction are never equilibrated in any reasonable time. We rather aim to study the conditions that guarantee mechanical stability. Such condition should be fulfilled whether the system is at thermal equilibrium or not, as long as it is stable on reasonably long time scales.

Note that since there is no energy involved in such system, the temperature only fixes the time unit. In what follows we impose for the average square velocity $\langle v^2 \rangle = 1$.

9.2 Effective potential

In the previous Chapters we studied the rigidity of amorphous solids by considering their vibrational modes. It is a priori problematic to use the same analysis to study the rigidity of contact networks such as the one of Fig.(9.1): the hard sphere potential is discontinuous, and the energy cannot be expanded as in Eq.(2.7). Nevertheless we shall go around this difficulty by deriving a smooth effective potential. The trick is to average on the fast fluctuations that take place on characteristic time smaller than some $t_1 \gg \tau_c$.

[22] If t_1 is too small some contacts can disappear, which can lead to the appearance of unstable modes when they are computed following the procedure bellow, even thought the system is stable

[23] Half of the particles have a diameter unity, the other half have a diameter 1.4.

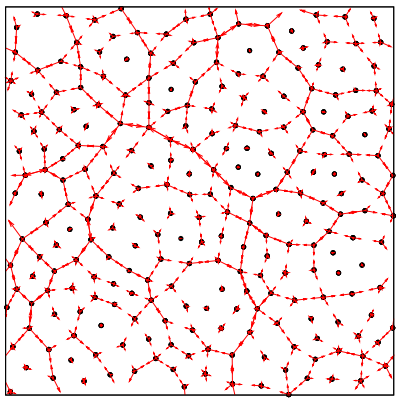

Fig. 9.1: Contact forces for $N = 256$, $\epsilon = 10^{-5}$ and $t_1 = 10^5$. Black points represent particles. Contact forces are sketch by arrows which start from the particle center, and whose length is proportional to the force amplitude. Note that the forces are balanced on every particle, as it must be the case on time scales where the structure is stable. For similar force networks see [101].

Consider two particles i and j in contact separated by a spacing $r_{ij} - r_i - r_j$ of typical value h, where r_i and r_j are the radii of the particles, and r_{ij} the distance between i and j. Such a spacing fluctuates in time between 0 (when the particles are colliding) and a few h, so that the instantaneous value of this spacing does not give much information about the contact ij. Nevertheless, if $r_{ij} - r_i - r_j$ is averaged on large time intervals $t_1 \gg \tau_c$, the spacing $h_{ij} \equiv \langle r_{ij} - r_i - r_j \rangle$ converges to a well defined value. We shall define the averaged position $\vec{R}^{av}(t)$ of particle i as:

$$\vec{R}_i^{av}(t) = \frac{1}{t_1} \int_{t-t_{\frac{1}{2}}}^{t+t_{\frac{1}{2}}} \vec{R}_i(t')dt'. \tag{9.2}$$

In what follows we estimate the contact force f_{ij} exchanged between two particles i and j whose spacing is $h_{ij} \approx ||\vec{R}_i^{av} - \vec{R}_j^{av}|| - r_i - r_j$. We furnish thermodynamic arguments, that apply both to Newtonian and Brownian dynamics. We start by considering a one-dimensional system with beads of diameter 1. The partition function Z is:

$$\mathcal{Z} = \prod_i \int_{h_i=0}^{h_i=\infty} dh_i e^{-ph_i} \tag{9.3}$$

where h_i is the spacing between particle i and $i+1$. If an external force dipole $p_i = -p_{i+1} \equiv p_1$ is applied on i and $i+1$, the partition function becomes:

$$\mathcal{Z} = \prod_{j \neq i} \int_{h_j=0}^{h_j=\infty} dh_j e^{-ph_j} \int_{h_i=0}^{h_i=\infty} dh_i e^{-(p+p_1)h_i} \tag{9.4}$$

From the partition function one can compute the average spacing $\langle h_i \rangle = \frac{1}{p+p_1}$. Since the contact force f_i in the contact $i, i+1$ is $f_i = p_1 + p$, one finds:

$$f = \frac{1}{h} \tag{9.5}$$

This result can be extended to an isostatic state of any spatial dimension. A particularity of the isostatic state is that the number of displacements degrees of freedom is precisely equal to the number of contact. Hence the configuration of the system can be described by the set of distances between particles in contact. If the system is at equilibrium in a meta-stable state where the contact forces field $|\mathbf{f}\rangle = \{f_{ij}\}$ is well-defined, the partition function can be written:

$$\mathcal{Z} = \prod_{\langle ij \rangle} \int_{h_{ij}=0}^{h_{ij}=\infty} dh_{ij} e^{-f_{ij}h_{ij}} \tag{9.6}$$

The argument valid for the one-dimensional line of particles is valid here, and one obtains $\langle h_{ij} \rangle = f_{ij}^{-1}$. This shows that in an isostatic assembly of colliding hard spheres, when the particles rattling are averaged, the hard sphere potential converge to an effective potential. At time $t \gg \tau_c$ the system behaves as an assembly of particles of positions $|\mathbf{R}^{av}\rangle = \{\vec{R}_i^{av}\}$ interacting with the potential $V_{ij}(r)$:

$$\begin{aligned} V_{ij}(r) &= \infty && \text{if} \quad r < r_i + r_j \\ V_{ij}(r) &\sim -ln(r - r_i - r_j) && \text{if } i \text{ and } j \text{ are in "contact"} \\ V_{ij}(r) &= 0 && \text{if } i \text{ and } j \text{ are not in "contact"} \end{aligned} \tag{9.7}$$

The relation force/distance is checked numerically in Fig. (9.2) at packing fraction close to ϕ_c. At such packing fractions the system is nearly isostatic, as we shall see in the next section. The exponent found is in very good agreement with Eq.(9.5). Fig. (9.2) also shows that for large t_1 the dispersion of the contact forces around their average value described by Eq.(9.5) is extremely small. This indicates that the only relevant parameter that characterizes the contact force amplitude is the spacing h, as predicted by Eq.(9.7).

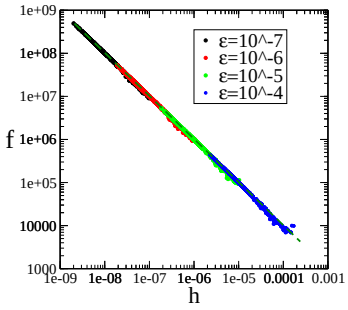

Fig. 9.2: Log-log plot of the contact force amplitude versus the spacing $h = r - r_i - r_j$ for various ϵ in a system of $N = 256$ particles. Each dot represents the pair of number $(f_{ij}, \langle h_{ij} \rangle)$ associated with the contact ij. Dots collapse on the dotted theoretical curve defined by Eq.(9.5).

When the coordination z increases, the h_{ij} are not independent variables anymore. We aim to evaluate the corrections to Eq.(9.7) when the system is at a finite distance from isostaticity, that is with $\delta z > 0$. Isolating the contact ij we may write the partition function as follows:

$$\mathcal{Z} = \int_{h_{ij}=0}^{h_{ij}=\infty} dh_{ij} e^{-f_{ij} h_{ij}} e^{-\mathcal{F}(h_{ij})} \tag{9.8}$$

where $\mathcal{F}(h_{ij})$ is the free energy of the entire system conditionally to the value of the spacing h_{ij}. It can be written as $\mathcal{F}(h_{ij}) = \mathcal{F}_0 + \Delta\mathcal{F}(h_{ij})$, where $\mathcal{F}_0$ does not depend on h_{ij}. To evaluate $\Delta\mathcal{F}(h_{ij})$, we consider as a zero order approximation that Eq.(9.7) is true. Then $\Delta\mathcal{F}(h_{ij})$ corresponds to the energy cost induced by a local strain of amplitude $h_{ij} - \langle h_{ij}\rangle = h_{ij} - f_{ij}^{-1}$ in an elastic system where particles interact with the potential (9.7). In Chapter 7 we computed the response to a local strain of amplitude e. The corresponding energy cost varies with the contact considered, and for small strain its amplitude follows $\delta E \sim \delta z B e^2$, where B is the bulk modulus. For an interaction given by Eq.(9.7), following chapter 7 one finds $B \sim \langle V''(r)\rangle \sim \langle (r - r_i - r_j)^{-2}\rangle \sim p^2$, so that $\Delta\mathcal{F}(h_{ij}) \sim \delta z p^2 (h_{ij} - f_{ij}^{-1})^2$. Thus we may write $\Delta\mathcal{F}(h_{ij}) \equiv \delta z C_{ij} f_{ij}^2 (h_{ij} - f_{ij}^{-1})^2$, where C_{ij} is positive, of order one, and can a priori depend on the contact considered. Using this expression in Eq.(9.8), and expanding Z to first order in δz, we can compute the corrections to $\langle h_{ij}\rangle$. One finds $\langle h_{ij}\rangle = \frac{1}{f_{ij}}(1 - 2C_{ij}\delta z)$, so that the force-displacement relation satisfies:

$$f_{ij} = \frac{1}{h_{ij}}(1 - 2C_{ij}\delta z) \tag{9.9}$$

This estimates the corrections to the potential of Eq.(9.7), which vanish when the system becomes isostatic. Thus, one non trivial consequence of these corrections is to weaken the force for a given inter-particle distance. To test this effect we compute numerically $C(\delta z) \equiv \langle f_{ij}\langle h_{ij}\rangle\rangle_{ij} - 1$, where $\langle\rangle_{ij}$ denotes the average over all contacts. The results are represented in Fig.(9.3). Small corrections are indeed found, which are in good agreement with our prediction $C(\delta z) \sim -\delta z$. In what follows we are mainly interested in scaling relations near ϕ_c, where corrections of the order δz do not matter. Thus we shall neglect them, and consider that the effective interaction is constant and given by Eq.(9.7).

Note that Eq.(9.7) can be recovered with a simple scaling argument. We may evaluate the collision frequency ν as $\nu \sim \frac{v}{h_{ij}}$, where v is the average amplitude velocity. The typical momentum exchanges during one collision is, taking $m = 1$ for the particles mass, $||\Delta\vec{P}|| \sim v$. Therefore following Eq.(9.1) we recover that the force follows $f_{ij} \sim \frac{\langle v^2\rangle}{h_{ij}} \sim \frac{1}{h_{ij}}$.

In what follows we shall only assume that the potential $V(r)$ is differentiable and that the force $f(h)$ scales as h^{-1}. The corrections estimated above in Eq.(9.9) do not affect these results, therefore we neglect them and

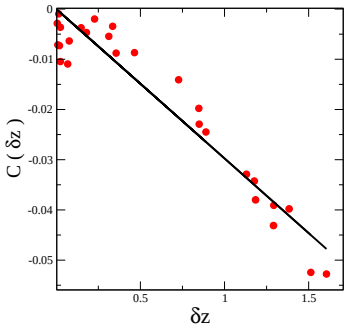

Fig. 9.3: Average correction $C(\delta z)$ as defined in the text *vs.* excess coordination δz for various ϕ. The line is a linear fit consistent with our predictions at small δz. Corrections are small, of the order of 3-4 percents when $\delta z = 1$.

use Eq.(9.7). As we shall see, this implies that ϕ_c acts also as a critical point in the liquid phase. In particular, Eq.(9.7) enables us to define the stiffness k_{ij} of a contact ij as:

$$k_{ij} = V''(r) \sim \frac{1}{(r - r_i - r_j)^2} \sim f_{ij}^2 \tag{9.10}$$

This allows to define a dynamical matrix and vibrational modes once the average particles positions and the contacts are known. As we discussed above, the potential of Eq.(9.7) has an entropic nature. Thus such vibrational modes describe the local curvatures of the entropic landscape of the system. In what follows we show that imposing the stability of these modes leads to constraint on the coordination of the force network. Then we discuss the elastic property of hard sphere glass, derive the elastic moduli and discuss the length scales that appear in the response of such systems. Finally we discuss how these modes may be related to the structural relaxation.

9.3 Stability of hard sphere systems

If the network of contact of a hard sphere system is weakly connected, we can apply the results of Chapters 3, 4 and 5. In particular, following Eq.(5.5) and Eq.(8.11) and the paragraph above it, we obtain that such system presents anomalous modes that appear at a frequency $\omega_{AM}^2 = A_1 B(p)\delta z^2 - A_2 p$, where $B(p)$ is the bulk modulus of the system. The bulk modulus scales as the average stiffness of the contacts, and following Eq.(9.10) we obtain $B(p) \sim p^2$. As we discussed in Chapter 5, a rigid system does not display

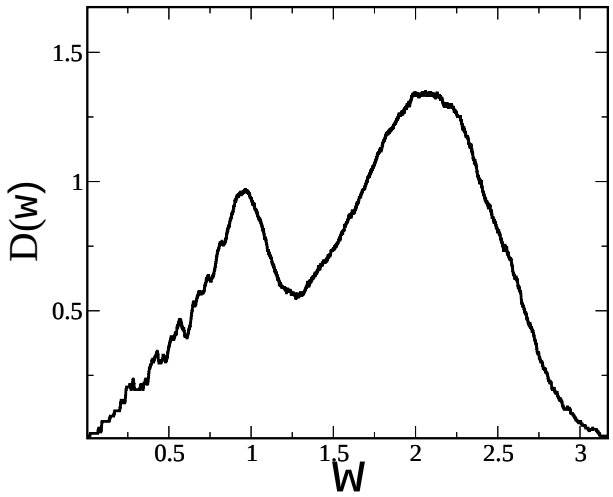

Fig. 9.4: Density of vibrational modes $D(\omega)$ versus frequency in a hexagonal crystal of 1024 particles for $\epsilon = 10^{-4}$. The frequencies are rescaled by ϵ^{-1}. The particle positions were average in time $t_1 = 4 \times 10^4$ to obtain $|\mathbf{R}^{av}\rangle$. Eq.(9.7) was used to compute the dynamical matrix, from which the vibrational frequencies were inferred.

**

unstable modes. Therefore we obtain that there must be a constant C_0 such that:

$$\delta z \geq C_0 p^{-\frac{1}{2}} \tag{9.11}$$

which is another realization of the extended Maxwell criterion we derived in Chapter 5. Note that near the jamming threshold, the typical inter-particle spacing ϵ goes as $\epsilon \sim \phi_c - \phi$ and thus $p \sim \epsilon^{-1} \sim (\phi_c - \phi)^{-1}$, as was already computed with different methods [79, 80]. Thus Eq.(9.11) indicates that δz must scale as $\delta z \sim (\phi_c - \phi)^{\beta}$ with $\beta \leq \frac{1}{2}$, as is the case for a soft sphere system *above* ϕ_c.

9.3.1 Stability of the hexagonal and the square crystals

To test Eq.(9.11) we perform tests on three different systems in two dimensions. We start by considering an hexagonal monodisperse crystal and a monodisperse square crystal. We consider these two systems at their maximum packing fraction where hard spheres are in contact. Then we reduce the particles diameter by a relative amount ϵ, and start the dynamics. According to Eq.(9.11) these systems must behave very differently. In the hexagonal

crystal, the coordination is 6, therefore $\delta z = 2 \gg p^{-\frac{1}{2}} \sim \epsilon^{\frac{1}{2}}$ for small ϵ. Therefore the condition of Eq.(9.11) is satisfied and we expect the system to be stable. On the other hand, in the square crystal, the number of first neighbors is 4, $\delta z = 0$, and the system cannot satisfy Eq.(9.11) without large structural rearrangements for any ϵ. Thus such a system cannot be rigid. These predictions are verified numerically. For small ϵ, the hexagonal crystal displays no structural changes, whereas the square crystal collapses rapidly. Such collapse leads to an hexagonal configuration, and is generated by the buckling of unstable modes. We show the corresponding displacements in Fig.(9.5). The instability can be also observed in the free energy landscape by computing the vibrational modes of the corresponding systems. Fig.(9.4) shows the density of vibrational modes of the hexagonal crystal, which varies linearly at low frequency, as expected for a two-dimensional crystal. No unstable mode are observed. In the square crystal, see Fig.(9.6), the density does not vanish at $\omega \to 0$, as expected for an isostatic system. Furthermore, we observe unstable modes, as implied by Eq.(9.11).

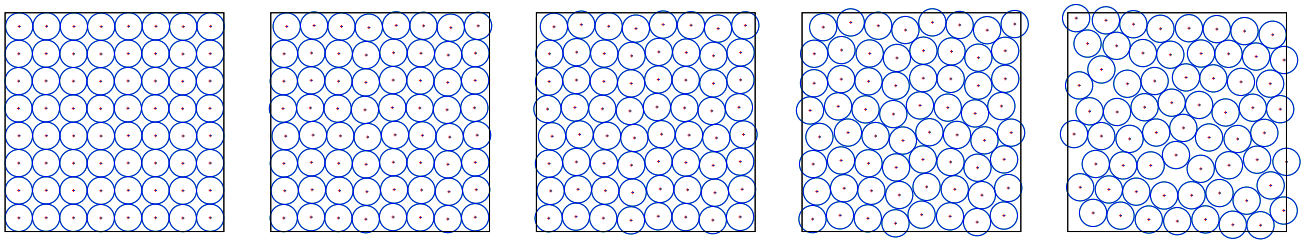

Fig. 9.5: Buckling of a square lattice.

9.3.2 Stability of hard sphere systems near the jamming threshold

We perform the same test using the polydisperse configurations of [50] at the jamming threshold. We study the structural stability of the system by computing the self correlation function $C(\vec{q},t) = \langle \exp[i\vec{q}\cdot(\vec{R}_j(t) - \vec{R}_j(0))]\rangle$ where the average is taken on every particle j. In what follows we consider wave vectors $\vec{q}$ of norm $\frac{\pi}{r_1}$, where r_1 is the radius of the smallest particle. Typical curves for different ϵ are represented in Fig.(9.7). For ϵ smaller that roughly $\epsilon = 0.05$, the system ages and dynamics is intermittent. There are

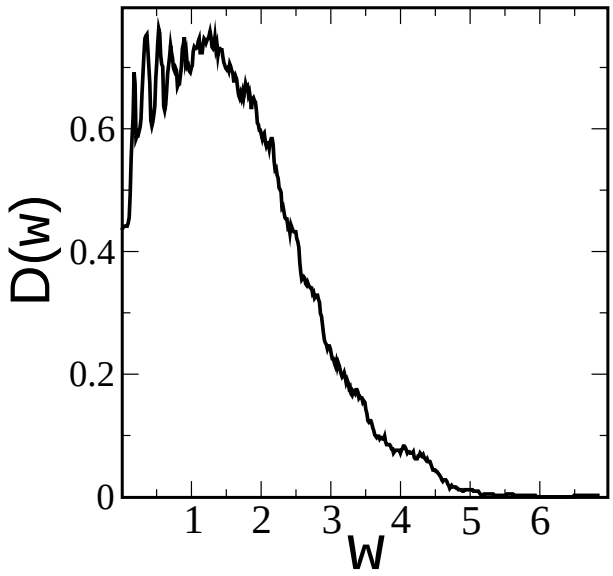

Fig. 9.6: Same caption as Fig.(9.4) for a N=1024 square crystal. 80 unstable modes were observed, lying in the frequencies range $\omega \in$

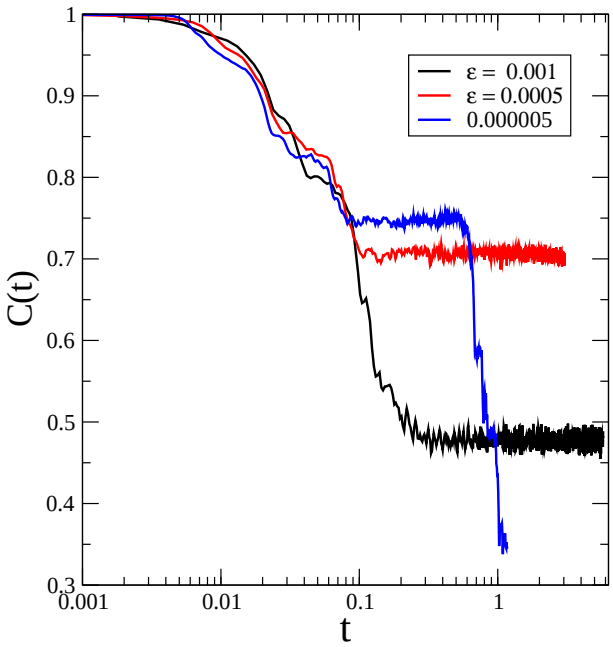

Fig. 9.7: Examples of $C(t) = C(q = 1, t)$ for $q = \pi/r_1$ versus real time for different ϵ in a $N = 256$ particles system. $t = 0$ corresponds to the initial time where random velocities are assigned to the particles.

long periods of time where the structure is stable, that appear as plateaus in the self correlation function. Such periods ends with crashes, where the self correlation function drops of a large amount in a very short time. These events must correspond to sudden collective rearrangements involving a large number of particles. During such crashes, we observe that the coordination increases, and that the pressure drops. We also observed that the correlation function of the force network $H(t) = \langle f_{ij}(\tau) f_{ij}(\tau + t) \rangle$ is constant during the plateau, and drops when a crash occurs.

Such crashes, or "earthquakes", were reported in other glassy systems. They occur when a Lennard-Jones liquid is rapidly quenched at temperature much lower than the glass transition [102]. In this case the crashes typically involve 100 particles. Such crashes were also observed in colloidal pastes using dynamic light scattering [103], and in the dielectric response of laponite [104]. It would be obviously interesting to understand what is the nature of these crashes, and what triggers the sudden collapse of apparently mechanically stable structure. These are fundamental questions of the glass transition. In the present Chapter we study the stable structures that appear before and after the crashes, when the system is quiet. Understanding why such structures are rigid is certainly a necessary step to find out why and how they can yield.

We define the quiet periods as the plateaus of $C(t)$. To study the rigidity of the dense hard sphere assemblies during these periods, we compute the vibrational modes. If the rattlers are removed [24], we find that, the system is mechanically stable: there are no unstable modes. The density of states for $\epsilon = 10^{-4}$ is shown in Fig.(9.8). The main difference with the other stable structure that we studied above— the hexagonal crystal— is the presence of a large excess of modes at low frequency. A large amount of modes are nearly unstable, as we expect near the jamming threshold where we must have $D(\omega) \to \omega^0$ when the frequency vanishes. Note that the density of states does not have a flat plateau as the harmonic soft spheres at the isostatic point, despite that the coordination network is the same when ϵ is small. The main difference between these systems is the stiffness disparity: in the amorphous hard sphere system, the disparity of the contact force leads to a disparity in the stiffness since $k \sim f^2$. If this disparity is removed by keeping the same contact network, but by imposing a constant stiffness in all contacts, we find that a flat plateau is recovered.

The absence of unstable modes enables us to test Eq.(9.11), that must be

[24] The rattlers do not participate to the rigidity of the structure. They can be identified as the distance with their neighbors is much larger than for the rest of the inter-particle distances, so that the frequency of the shocks they have with their neighbors is much smaller than the other particles.

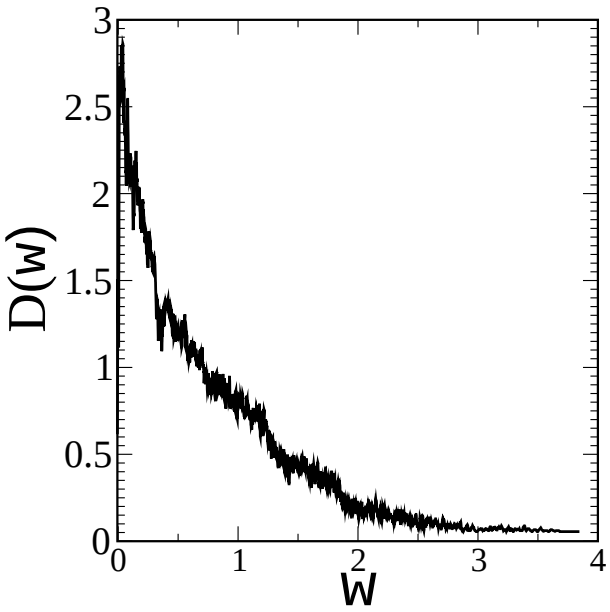

Fig. 9.8: Density of vibrational modes $D(\omega)$ versus frequency in a amorphous hard sphere glass of 256 particles for $\epsilon = 10^{-4}$. The frequencies are rescaled by ϵ^{-1}. $D(\omega)$ was computed in a quiet period preceding the first crash. No unstable modes were observed.

satisfied as the system is stable.To check this relation we computed numerically both the coordination and the pressure for various packing fractions, and for various stable periods that appear along the aging regime. As shown in Fig.(9.9), the data are consistent with an *equality* of the inequality (9.11). This suggests that a hard sphere glass is only *marginally* stable, as it is the case for soft spheres slowly decompressed toward ϕ_c.

9.4 Elastic property of the hard sphere glass

We now use this approach to derive the elastic behavior of the hard sphere glass near ϕ_c. The results of Chapter 7 apply. In this system the elastic moduli have purely entropic causes: they describe how the number of configurations is reduced under a global strain. As we already discussed, the scaling of the bulk modulus B is:

$$B \sim p^2 \sim (\phi_c - \phi)^{-2} \tag{9.12}$$

In repulsive systems near ϕ_c the shear modulus scales as the bulk modulus times the excess coordination. Thus the extension of Eq.(7.33) to non

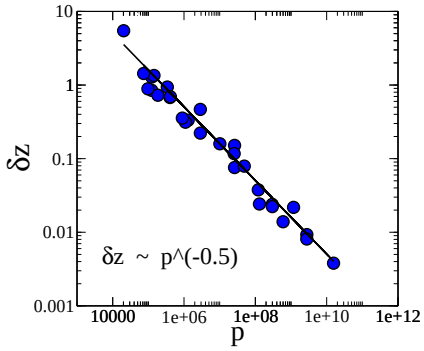

Fig. 9.9: Log-log plot of δz versus the average contact force $\langle f \rangle \sim p$. The data were obtained for different ϵ and different time periods. The black line corresponds to the equality of the inequality (9.11).

harmonic contacts yields:

$$G \sim B\delta z \sim \delta z p^2 \tag{9.13}$$

If one assumes that the glass is marginally stable, that is to say that the system lies on the bound of inequality (9.11), one obtains:

$$G \sim p^{3/2} \tag{9.14}$$

As in practice such glass is not at equilibrium near ϕ_c, the hypothesis of marginal stability certainly depends on the preparation of the system. It is reasonable to think, following the analogy with the soft spheres, that if the system is slowly compacted toward ϕ_c the dynamics would not bring the system far away from marginal stability. Thus we expect $\delta z \sim p^{-\frac{1}{2}}$ to be a good approximation and the shear modulus G not to differ to much from the scaling law of Eq.(9.14).

Furthermore, this system present excess modes: there are the anomalous modes we discussed in the previous Chapters. Near the jamming threshold, when the pressure diverges, we expect $D(\omega) \to \omega^0$. When ϕ decreases, the density of states is characterized by some frequency $\omega^* \sim B^{\frac{1}{2}}(p)\delta z \sim p^{-\frac{1}{2}}\delta z$. Consequently a hard sphere glass is characterized the length $l^* \sim \delta z^{-1}$ which characterizes the anomalous modes present at low-frequency.

9.5 Discussion

In this Chapter we studied some properties of the hard-sphere glass near the close packing at ϕ_c. We argued that the mechanical stability at time

$t \ll \tau_\alpha$ relies on the emergence of a rigid force network. The key point was to compute the properties of the force networks by averaging on fast fluctuations, the particles rattling. This allows one to derive an effective potential, exact at the isostatic limit, and to define rigorously the vibrational modes in hard sphere systems. They characterize the entropic landscape expansion around a configuration. This allows one to recover the main results valid for soft sphere systems near the jamming threshold. In particular the stability of the vibrational modes imposes a bound on the coordination of the contact force network. More generally, we showed that ϕ_c acts as a critical point: the elastic moduli and the coordination z scale as $\phi \to \phi_c$. The system responses are characterized by l^* that diverges at ϕ_c.

As critical points and diverging length scales are in general associated with diverging time scales, the presence of a critical behavior at ϕ_c may be the cause for the glass transition observed in hard sphere systems. In particular, our work suggests that the slowing-down of the dynamics is due to the appearance of an *extended* rigid network of interactions, rather than to any local properties such as the tightening of cages. This suggests an alternative perspective for the relaxation in super-cooled liquid, at equilibrium or in an aging regime such as the one of Fig.(9.7). The particularity of the systems that crash, such as the square crystal or the amorphous state near isostaticity, is the large amount of modes around zero frequency. As we discussed in previous Chapters such modes are very different from plane waves, as they have rapid spatial fluctuations, which makes them sensitive to the applied stress term responsible for buckling. Furthermore they can be localized on length scales larger than some l^* without changing in their frequency. These modes indicate a nearby elastic instability, and may well play a role in the rigidity loss. Furthermore, whatever rearrangements occur when the system relaxes, suddenly or not, the corresponding displacements must be similar to the soft modes, as there are, rougly speaking, the only modes where particles can avoid to inter-penetrate. Thus it is tempting to associate the structural relaxation with the structural buckling of the weakest frequency modes, that could be induced for example by pressure or coordination fluctuations. If so, we expect the relaxation to occur on length scales of order $l^* \sim p^{\frac{1}{2}}$, since for smaller subsystems the frequency of the modes increases, and the elastic instability goes away. Thus our work suggests an original perspective on the possible cause of the heterogeneous dynamics that occur in super-cooled liquids [21, 20]. Many models of the glass transition lead to heterogeneous dynamics [22]. In most of them, a local rule describes the motion of one or few particles, and leads to cooperative dynamics. Our work suggests an alternative view: the rule that allows the particle motion is *itself* non local, for the simple reason that rigidity is not global property, as was already

understood by Maxwell more than one century ago.

10. Conclusion

10.1 Summary

The elastic properties of an assembly of short-range, repulsive particles display critical behaviors near the jamming transition. In particular the elastic moduli, the frequency of the excess-modes, and the coordination number scale. In the first part of this thesis we derived the corresponding exponents. Then we showed how some of these results can apply to various systems, such as silica glass or colloidal particles.

The starting point is the following variational argument: if a rigid system of coordination z is cut in sufficiently small subsystems of size l, these subsystems are not rigid. This is true as long as the subsystems size is smaller than some length $l^* \sim \delta z^{-1}$ that diverges at the jamming threshold. Smaller subsystems contain modes of zero frequency, the soft modes. From these modes one can build what we called the "anomalous modes" which have a frequency of order $1/l$ in the original system. This gives for the dependence of the onset of excess-modes $\omega^* \sim \delta z$, as observed numerically. The system can be described as a continuous elastic medium only for $\omega \leq \omega^*$. At larger frequencies, it behaves as an isostatic state.

Then, by arguing that the soft modes have large transverse relative displacements, we showed that the anomalous modes were much more sensitive to the applied stress than the conventional acoustic modes. This has two direct consequences: in a purely repulsive system, anomalous modes can appear at frequencies much smaller than ω^*. Furthermore, in rigid solids the coordination number must be sufficiently large to guarantee the stability of the anomalous modes. We find $\delta z \geq C_0(p/B)^{\frac{1}{2}} \sim (\phi - \phi_c)^{\frac{1}{2}}$, where p is the pressure and B the bulk modulus. This generalizes the Maxwell criterion for rigidity $\delta z \geq 0$ valid when applied stress is absent. It follows that, if the jamming threshold ϕ_c is reached adiabatically, the pair correlation $g(r)$ measured at ϕ_c must display a divergence $g(r) \sim \frac{1}{(r-1)^\gamma}$ with $\gamma \geq \frac{1}{2}$. This divergence is the vestige of the excess contacts δz necessary to maintain the rigidity of the structure at larger ϕ.

These arguments furnish the scalings of both the onset of excess modes and the coordination number, but they do not enable us to compute the elas-

tic moduli. To do so, we used the fact that the linear equation that defines the soft modes is the dual of the force balance equation. This allows us to derive an original formalism relating the response to a strain to the contact force fields that satisfy force balance on each particle. This enables us to compute the shear and the bulk modulus. We found that for a repulsive system the ratio G/B vanishes at the jamming transition. We also obtain exact results on the response to a local strain. The energy cost of such deformation vanishes at the transition as the shear modulus. At the transition, this response extends in the entire system. Therefore this is also true for the soft modes that appear when one contact is cut in an isostatic state, an assumption that was essential to compute the frequency of the anomalous modes.

In a second part we studied the applications of these concepts to glasses, granular matter and colloids. All glasses have an excess of low-frequency modes, the boson peak. It is especially strong in silica, one of the best glass former. In silica the strongest interactions are those responsible for the rigidity of the SiO_4 tetrahedra. If the weaker interactions, such as Van de Waals, are neglected, the tetrahedral network obtained is isostatic. Following our argument, such a network must have a constant density of states at low frequency, as it is observed numerically. When the weaker interactions are turned on, the anomalous modes responsible for the plateau shift to higher frequencies. This improved variational argument predict the appearance of a plateau in the density of states. Such plateau is indeed observed in the simulations of silica, and appears around 1 THz, the boson peak frequency. As the key parameter of our description is coordination rather than disorder, this argument also justifies the similarity between the density of states of silica and the one of the corresponding crystals, which also present a plateau at roughly the same frequency. This is certainly a strong point, as most of the existing boson peak theories are based on disorder only, and cannot explain the excess modes that show up in these crystals. Finally we proposed to extend these ideas to Lennard-Jones systems.

Generally critical points display similar behavior on both sides of the transition. We argued that it is also the case for the jamming transition. We showed that when a hard sphere liquid approaches the jamming transition, the contact force network, that now characterizes how particles exchange momentum, is very similar to the one of elastic spheres above ϕ_c. The key point was to show that when the dynamics of an isostatic system is averaged on short time scales, the interactions among particles can be described by a logarithmic effective potential. We evaluated the corrections of this potential when the coordination increases. This allows to compute the normal modes of such systems. As a consequence the results of elastic spheres solids apply

to hard sphere liquids. In particular the extended Maxwell criterion must be satisfied near ϕ_c, when the system is rigid on short time scales. This approach also yields the elastic moduli that characterize the system for $t \ll \tau_\alpha$, the relaxation time. This description supports that the relaxation in supercooled liquid should not be described in terms of the motion of individual particles, even if models with such local rules can lead to non-trivial cooperative dynamics. Rather, it suggests that the elementary motions to consider are themselves collective, and are related to the anomalous modes introduced above.

10.2 Perspectives

10.2.1 Low temperature glass properties

The transport properties of glasses, but also of granular matter [105, 106], are not understood [3]. In particular there is a plateau in the thermal conductivity around 10 K, temperatures at which the heat transport can be dramatically smaller than in the crystal phase [5]. These temperatures correspond to frequencies in the Thz range, where the boson peak appears. In repulsive, short-range systems we showed that the cause for the boson peak is the following: above some frequency ω^* the system behaves as an isostatic state, whose density of states is much larger than the one of the crystal at low frequency. We argued that these ideas are more general and apply for example to silica and granular matter. This suggests that the transport properties of glasses in the Thz range correspond to those of a system at the jamming threshold. Then the question is to understand how the anomalous modes we introduced contribute to the transport. Accordingly, it is necessary to compute their spatial power spectrum $E(q)$. For a normal mode of displacement $\{\delta\vec{R}_i\}$, $E(q)$ is defined as:

$$E(q) = \frac{1}{q^2}\sum_{j,l}(\vec{q}\cdot\delta\vec{R}_j)(\vec{q}\cdot\delta\vec{R}_l)e^{i\vec{q}\cdot(\vec{R}_j-\vec{R}_l)} \tag{10.1}$$

$$\sim \frac{N}{q^2}\int dr^d\langle(\vec{q}\cdot\delta\vec{R}(r))(\vec{q}\cdot\delta\vec{R}(0))\rangle e^{i\vec{q}\cdot\vec{r}} \tag{10.2}$$

For an acoustic mode of frequency ω, $E(q)$ presents a peak for $q = c^{-1}\omega$, where c is the longitudinal velocity of sound. If this peak has a finite width Δq, the scattering length of the acoustic mode follows $l \approx \Delta q^{-1}$. This length enters in the computation of the thermal conductivity: the contribution of a mode to the heat transport goes as $l\cdot c$ [1]. The empirical data suggests that at frequency smaller than the boson peak, the acoustic modes have a scattering

length $l \approx 150\lambda$, where λ is the wavelength of the corresponding mode [3]. We expect that the anomalous modes that appear at higher frequencies will transport much less than that, because their spatial correlations $\langle \delta\vec{R}(0) \cdot \delta\vec{R}(x) \rangle$, which enter in Eq.(10.1), are presumably very small. This quantity is similar to the spatial correlations of the soft modes. If these correlations were zero, both the soft modes and the anomalous modes would be equally distributed on all wave vectors. Then the anomalous modes would have a purely diffusive behavior, corresponding to a scattering length $l = 1$ particle size. Thus they would almost not contribute to the transport at all. It is apparent from Fig.(3.1) that the spatial correlations of the soft modes are indeed small. This observation may furnish an explanation for the observed bad quality of the transport at the boson peak frequency. It is also consistent with the empirical observation that the boson peak frequency depends only weakly on wave vectors q if at all [107], indicating that the excess-modes have a wide distribution over the wave vectors q. For a quantitative discussion it would be of great interest to derive the spatial correlations induced by the soft mode equation (2.3). Furthermore, other interesting effects could in principle affect the transport. In particular, our derivation of the anomalous modes does not preclude the presence of underlying acoustic modes, that could hybridize with the anomalous modes, and enhance the transport.

At lower temperature, the properties of glasses are still a challenge to theory. The specific heat has a nearly linear dependence with temperature, and the thermal conductivity varies quadratically. This is in general interpreted by the presence of two-levels systems: atoms or groups of atoms can switch between two configurations by tunnel effect. This model is phenomenological and there is no consensus on what these two level systems may be. It has been often argued that the excess-modes of the boson peak are good candidates to form two-levels systems, see e.g. [35, 36]. As these modes are soft, non-harmonic terms are important, which could lead to the canonical form of 2-levels systems: two wells separated by a potential barrier. If it is so, our interpretation of the boson peak suggests that the 2-level systems are rather extended, plausibly on the length l^*, and that the displacement of each particle could be much smaller than a particle size. It should be possible in principle to test this possibility. As we discussed earlier, the pressure has two opposite effects on the anomalous modes: on the one hand it increases the coordination, and on the other hand the applied stress term lowers their frequency. We expect that in some systems, the destabilizing effect of pressure dominates. This can even lead to an elastic instability where anomalous modes become unstable, and where the configuration of the system changes. Such elastic instability might occur for example in silica glass at high pressure

[108]. If the density of anomalous modes is increased at very low temperature by tuning the pressure, the density of two-levels systems should be enhanced too. The latter could be directly checked by specific heat measurements.

10.2.2 The glass transition

We discuss the following possibility: in fragile glass former a fast slow down of the dynamics occurs close to the temperature at which the system manages to form an extended rigid network. We first study the role of temperature on the anomalous modes in the glass phase. Then we propose a microscopic distinction between strong and fragile glasses [26], and we discuss a possible structural relaxation process in the fragile case.

Role of temperature

Following the results of Chapter 8, we can write for the onset of the anomalous modes at zero temperature $\omega_{AM}^2 = A_1 B \delta z^2 - \mathcal{H}(p) + \zeta$, where B is the bulk modulus, δz is the effective excess-coordination number, ζ quantifies the effect of the weak interactions on the anomalous modes, and A_1 is a numerical constant. $\mathcal{H}(p)$ is the correction induced by initial stress term. For soft spheres near the jamming threshold where the distances between particles in contact are similar we found $\mathcal{H} \approx A_1 p \sim \langle f \rangle$, where $\langle f \rangle$ is the average contact force. In a system where the distance r between interacting particles can vary, following Eq.(2.9) we have $\mathcal{H} \sim \langle f/r \rangle$. In the present qualitative discussion we shall neglect these corrections and consider $\mathcal{H} \approx A_1 p$.

If the system is heated at constant pressure, we may extend this equation and write:

$$\omega_{AM}^2(T) = A_1 B(T) \delta z(T)^2 - A_2 p + \zeta(T) \tag{10.3}$$

In most glasses, when the temperature increases at constant pressure, the volume grows, and B decreases. As the typical inter-particle distances grows, we expect both the effective coordination and the effect of the weak interactions to diminish. Thus all the positive terms in Eq.(10.3) decrease. On the other hand, the pressure is constant. Hence $\omega_{AM}^2(T)$ decreases, as is indeed observed in most glasses. Eventually $\omega_{AM}^2(T)$ reaches zero frequency, as it has been observed numerically [32] and empirically [28]. We shall denote the temperature at which rigidity is lost T_r. Note that in few materials, such as silica, the volume *decreases* and the bulk modulus grows with temperature.

[26] The relaxation time of strong glass as an Arrhenius dependence with temperature, and grows faster in a fragile glass.

In this case Eq.(10.3) predicts that the boson peak shifts to higher frequency, as observed empirically, see e.g. [46].

Note that when $T > T_r$, there are continuously unstable modes. Interestingly, such a temperature also exists in mean field spin models [109], which were proposed as possible scenarios of the glass transition [110]. The dynamics of these models at higher temperature is exactly described by mode coupling equations [111]. According to this analogy $T_r = T_{MCT}$.

Fragile and strong glasses

In order to discuss the distinction between strong and fragile glasses, we introduce a second characteristic temperature T_a. T_a corresponds to the typical energy activation of the structural relaxation processes that do not depend on the rigidity of the structure. In particular we think about local rearrangements, such as the displacements of coordination defects in strongly covalent networks [112]. The glass transition must occur around the smallest of the two temperature T_a and T_r, since by definition the structure relaxes easily at higher temperature. Hence if $T_a \gg T_r$ the glass transition occurs in the vicinity of T_r. As the curvatures of the energy landscape dramatically evolves with temperature near T_r, it is natural to expect the corresponding dynamics to be super-activated. This supports the following scenario: glasses with $T_a \gg T_r$ are fragile. On the other hand, if $T_a \ll T_r$, the glass transition takes place in the vicinity of T_a. As the local structure does not display any important changes near the glass transition, we expect such glasses to be strong. A similar discussion in terms of energy landscape is presented in [114].

This scenario implies that strong glasses are whether (i) system with a large coordination, where anomalous modes frequencies are high. This is coherent with the empirical fact that strongly connected covalent networks are strong. (ii) anomalous systems which have a strong boson peak, but where an increase of temperature does not lower much the boson peak frequency, or even stabilizes the system like in silica. In these two cases we expect the rigidity of the covalent network not to do play any role at the glass transition. This is supported by simulations that indicate that the structural relaxation is purely local in silica glass [112], and that the covalent network exists until 8000 K at our pressure [113], which is much larger than the glass transition temperature. Note the elastic anomaly of silica disappears at high pressure, which suggests that this glass might become fragile when the pressure increases [115].

According to the present point of view, fragile glasses must display anomalous modes that approach zero frequency around the glass transition. A pos-

sible test could be done by considering particles with gaussian potentials. It was shown theoretically and numerically [42, 43, 44] that such systems at infinite temperature display excess-modes. These modes become stable above a finite density. Above this density this scenario predicts that the glass is strong, as the stable excess-modes at infinite temperature will stay stable as the temperature is lowered. At smaller density we expect a fragile behavior to occur, as it is the case for soft spheres [117].

Heterogeneous relaxation in fragile glass

According to the present scenario, in a fragile glass at T_r the anomalous modes are characterized by a finite length scale l^*. According to Eq.(10.3) $l^* \sim \delta z^{-1}$ does not diverge since the initial stress term, and the pressure, have a finite value. Thus we also expect the shear modulus at T_r to have a finite value, as it is the case for the marginally rigid system of [50]. As soon as $T \geq T_r$, the system displays unstable normal modes. Near T_r, only the modes with characteristic length l^* are unstable. The anomalous modes confined on subsystems smaller than l^* have higher, non-zero frequencies. Thus the collapse of unstable modes at T_r involve rearrangements on length scale of the order of l^*, but not smaller. As the temperature increases, modes with smaller characteristic lengths become unstable, and rearrangements can occur on shorter length scales. Hence this model predicts the presence of a growing length scale $l(T)$ that converges toward l^* when the temperature decreases toward T_r. Growing dynamical length scales were observed numerically as the temperature decreases, see e.g.[116] and ref. therein. The curve $l(T)$ could be measured by pinning the particles at the boundary of subsystems of size l, and by considering the dependence of the structural relaxation time with temperature for given l. Note that a similar test was already proposed in [119] to test dynamical length scales.

When the temperature decreases below T_r, we expect activated events to relax the glass structure. This is plausibely enhanced by the neighboring elastic instability, and by the large amount of nearly unstable modes. Finding the relaxation process of such weak structures is a necessary next step. Such models may lead to the prediction for the super-activated dependence of the relaxation time when the temperatures decreases and the elastic instability goes away. Once again, rigidity is not a local criterion, but demands the existence of a relation between coordination number and pressure on length scale l^*. Hence it is not excluded that the fluctuations of quantities such as pressure and coordination on such distances may trigger relaxations.

10.2.3 Granular matter

There are open questions on the way force propagates in amorphous systems such as granular matter. At large distances, a granular pack should behave as a continuous elastic medium, at least in a small linear regime. As we discussed in Chapter 2, the situation is different in anisotropic isostatic systems, whose elastic behaviors can be described by hyperbolic equations. Nevertheless, there is as yet no description of force propagation in *isotropic* isostatic systems. More importantly, we do not know how force propagation evolves toward a normal continuous elastic behavior when the coordination number increases, or when friction is present. It was proposed in [57] that in anisotropic system, normal elasticity is recovered above l^*, the distance at which soft modes disappear. Our derivation of the anomalous modes suggests that the transition toward continuous elasticity could occur at distances shorter than l^*. In particular for a force *dipole*, or a local strain, as we discussed in Chapter 8 it is plausible that the characteristic transverse length $l_t \sim \delta z^{-\frac{1}{2}}$ characterizes the cross-over isostaticity/elasticity. Note that this does not preclude that the response to a force mono-pole has a different cross-over. It would be of much interest to investigate these subtle properties. It could be done in simulations of elastic spheres near the jamming threshold. Experimentally such tests would require to find efficient ways to modulate the coordination and the distance from isostaticity [118].

The length scales we are talking about obviously depend on the system considered, but we expect them to be typically of the order of a few tens of particle sizes. Thus they may not affect for example the building of sand castles, as a continuum description is expected to be valid for macroscopic objects. Nevertheless the properties of an assembly of grains at such length scales might play a crucial role in the rheology of these systems, both in the solid and in the dense liquid phases. Similar length scales were observed in the spatial correlations of dense granular flows on a slope, and are probably related to the surprising dependence of the thickness h of the flow with the angle of the slopeθ [17]. Bending a layer of sand toward its avalanche angle is equivalent to imposing a shear. Thus understanding how shear affects the anomalous modes might shed light on this problem. The presence of a fixed boundary at a distance h of the free surface certainly increases the anomalous modes frequency. Making the most simple assumptions that (i) the shear decreases linearly the characteristic energy of the anomalous modes and (ii) the presence of of a fixed boundary is equivalent to an increase of coordination of order $1/h$ gives $\omega_{AM}^2 \sim A_1 B(\delta z + 1/h) - A_3 \sin\theta$. Imposing $\omega_{AM} = 0$ yields a dependence of h of the form $h(\theta) \sim \frac{1}{\theta - \theta_0}$, which is not inconsistent with the empirical data.

Another interesting property of granular matter is compaction: if a container of sand is tapped from below, the density of the system increases. The process displays many glassy features: the dynamics becomes very slow with time, aging and memory effect are observed [16]. When a granular pack is tapped weakly enough to avoid fluidization, there are two main causes for the irreversible events that lead to compaction. One the one hand, contacts on the Coulomb cone can slide. On the other hand, the destabilizing effect of the pressure wave can cause the buckling of anomalous modes. This structural buckling increases the coordination and the packing fraction, as sketched in Fig.(6.2). Hence the structure becomes more and more stable, and the compaction dynamics slows down. As the coordination rises the length scale l^* decreases. This may be possible to test this presiction using X-ray microtomography.

11. Acknowledgments

I am happy to thank Jean-Philippe Bouchaud, Carolina Brito, Sidney Nagel, Leonardo Silbert and Thomas Witten who contributed to this work.

References

[1] *Amorphous solids, Low temperature properties*, edited by W.A. Phillips (Springer, Berlin, 1981)

[2] P.W. Anderson, B.I. Halperin and C.M. Varma, the philosophical magazine, **8**, 25

[3] J.J. Freeman, A.C. Anderson, Phys.Rev.B, **34**, 5684 (1986)

[4] S. Ludwig, C. Enss, P. Strhlow, S Hunklinger, Phys. Rev. Lett.,**88**, 075501 (2002)

[5] R. Berman, Proc. R. Soc. A **280** 90 (1950)

[6] A. Tanguy, J.P. Wittmer, F. Leonforte and J.L. Barrat. Phys. Rev. B,**66**,174205 (2002)

[7] J.P. Wittmer, A. Tanguy, J.L. Barrat and L. Lewis, Europhys. Lett.,**57**, 423 (2002)

[8] F. Leonforte, A. Tanguy, J. P. Wittmer, and J.-L. Barrat, Phys. Rev. B,**70**, 014203 (2004)

[9] F. Leonforte, R. Boissiere, A. Tanguy, J. P. Wittmer, and J.-L. Barrat, cond-mat 0505610

[10] R. Brockbank, J.M. Huntley and R.C. Ball, J. Physque II, **7**, 1521 (1997); and ref. therein

[11] J.P. Wittmer, P. Claudin, M.E. Cates and J.P. Bouchaud, Nature, **382**, 336 (1996)

[12] M.E. Cates, J.P. Wittmer, J.P. Bouchaud, P. Claudin, Phys. Rev Lett., **81**, 1841 (1998)

[13] E.Clement, G.Reydellet, L. Vanel, D.W. Howell, J.Geng, R.P. Behringer, *XIII international congress on rheology, Cambridge (UK)*, Vol. **2** (British Society of Rheology, Glasgow, 2000) p.426

[14] G. Reydellet and E. Clement. Phys. Rev. Lett.,**86**, 3308 (2001) and refs. therein.

[15] J.-P. Bouchaud, P. Claudin, D. Levine, and M. Otto, Eur. Phys. J. E,**4**, 451-457 (2001)

[16] Nowak, E. R., Knight, J. B., Ben-Naim, E., Jaeger, H. M., Nagel, S. R., Phys. Rev. E, **57**, 1971?1982 (1998); Philippe, P. Thesis, Univ. Rennes I (2002); Josserand, C., Tkachenko, A. V., Mueth, D. M., Jaeger, H. M. Memory effects in granular materials. Phys. Rev. Lett. **85**, 3632?3635

[17] O. Pouliquen, Phys.Rev.Lett. /bf93, 248001 (2004)

[18] A.J. Liu and S. Nagel, Nature, **396** N6706, 21 (1998)

[19] C.A. Angell, in *Relaxation in complex systems*, eds. K.L. Ngai and G.B. Wright, Washington DC (1985)

[20] E. Weeks and D. A. Weitz, Chemical Physics **284** 361-367

[21] see e.g. M. D. Ediger, Ann. Rev. Phys. Chem. **51**, 99 (2000)

[22] C. Toninelli, M. Wyart, L. Berthier, G. Biroli, J.-P. Bouchaud, Phys. Rev. E **71**, 041505 (2005)

[23] W. Kob and H C Anderson, Phys. Rev. E, **48** 4364 (1993)

[24] Maxwell, J.C., Philos. Mag., **27**, 294-299 (1864)

[25] C. Kittel, Introduction to solid state physics, edited by John Willey and son, Inc.

[26] S.N. Taraskin and S. R. Elliot, Phys. Rev. B,**59**, 8572 (1999)

[27] A. P. Sokolov, U. Buchenau, W. Steffen, B. Frick, and A. Wischnewski, Phys. Rev. B,**52**, 9815 (1995)

[28] N. J. Tao, G. Li, X. Chen, W. M. Du, and H. Z. Cummins, Phys. Rev. A,**44**, 6665 (1991)

[29] D. Engberg, A. Wischnewski, U. Buchenau, L. Brjesson, A. J. Dianoux, A. P. Sokolov, and L. M. Torell, Phys. Rev. B, **59**, 4053 (1999)

[30] A.I. Chumakov, I. Sergueev, U. Van Burck, W. Schirmacher, T. Asthalter, R. Ruffer, O. Leupold, and W. Petry, Phys. Rev. Lett., **92**, 245508, (2004)

[31] A.Wischnewski,U. Buchenau, A.J. Dianoux, W.A. Kamitakahara, J.L. Zarettky, phys. rev. B,**57**, 2663-2666 (1998)

[32] J. Horbach, W. Kob and K. Binder, Eur. Phys. J. B **19**, 531-543

[33] G. Parisi, Eur. Phys. J.E., **9** 213-218 (2002)

[34] Grigera TS, Cavagna A, Giardina I, Parisi G, Phys. Rev. Lett.,**88**, 055502 (2002)

[35] U. Buchenau, Yu.M Galperin, V.L. Gurevich, H.R Scober, Phys. Rev. B **43**, 5039

[36] S.Alexander, Phys. Rep.,**296**, 65 (1998)

[37] V.G. Karpov, M.I Klinger, F.N. Ignatiev, Zh. Sov. Phys. JETP **57**, 439

[38] V.L. Gurevich, D.A. Parshin, J. Pelous, H.R. Schober, Phys. Rev. B **48**, 16318

[39] W. Schirmacher, G. Diezmann and C. Ganter, Phys. Rev. Lett.,**81**, 136 (1998)

[40] S.N. Taraskin, Y.L. Loh, G. Natarajan, and S.R. Elliot, Phys. Rev. Lett.,**86**, 1255 (2001)

[41] Maurer E, W. Schirmacher Jour. of low temp. Phys. 137 (3-4): 453-470 NOV 2004

[42] S. Ciliberti, T.S. Grigera, V. Martin-Mayor, G.Parisi and P. Verrocchio, AIP Conference Proceedings 708, 565 (2004)

[43] T.S. Grigera, V. Martin-Mayor, G.Parisi and P. Verrocchio, Phys, Rev. Lett., **8** 085502 (2001)

[44] M. Mezard, G.Parisi, A. Zee, Nuclear Phys. B, **559** 689-701 (1999)

[45] W. Gotze and M.R. Mayr, Phys. Rev. E, **61**, 587 (2000)

[46] T. Nakayama, Rep. Prog. Phys. **65** 1195-1242 (2002)

[47] A.J. Leadbetter, J. Chem. Phys. **51** 779 (1969)

[48] D. Caplin, G. Gruener and J.B. Dunlop, Phys. Rev. Lett. **30** 1138 (1973)

[49] Bilar and W.A. Phillips Philos. Mag. **32** 113 (1975)

[50] C.S O'Hern, L.E Silbert, A. J. Liu and S.R. Nagel, Phys. Rev. E, **68**, 011306 (2003)

[51] C.S. O'Hern, S.A. Langer, A.J. Liu and S.R. Nagel, Phys. Rev.Lett., **88**, 075507 (2002).

[52] D.J. Durian, Phys. Rev. Lett., **75**, 4780 (1995)

[53] L.E. Silbert,D. Ertas, G.S. Grest, T.C. Hasley, and D. Levine, Phys. rev. E, **65**, 031304 (2002)

[54] T.G. Mason, J. Bibette and D.A. Weitz, *Phys. Rev. Lett.*, **75**, 2051 (1995)

[55] J.C. Phillips, Jour. of Non-Crystal. solid,**43**, 37-77 (1981)

[56] M.F. Thorpe, J.Non-Crys. Solids, **57**, 355 (1983)

[57] A.V. Tkachenko and T.A Witten, Phys. Rev. E, **60**, 687 (1999);

[58] A.V. Tkachenko and T.A Witten, Phys. Rev. E, **62**, 2510, (2000)

[59] D.A. Head, A.V. Tkachenko and T.A Witten, European Physical Journal E,**6** 99-105 (2001)

[60] C.F. Moukarzel, Phys. Rev. Lett., **81**, 1634 (1998)

[61] P.M. Duxbury, D.J. Jacobs, M.F. Thorpe and C. Moukarzel,Phys. Rev E, **59** 2084 (1999)

[62] N.W. Aschcroft and N.D. Mermin, Solid state physic *HWR International*.

[63] Y.Cai and M.F. Thorpe, Phys. Rev. B, **40** 10535

[64] M. Tatsumisago, B.L. Halfpapa, J.L. Green, S.M. Lindsay and C.A. Angell, Phys. Rev. Lett., **64**, 1549 (1990)

[65] J-N Roux, Phys. Rev. E, **61**, 6802 (2000)

[66] L. Silbert, A.J. Liu and S. Nagel, *cond-ma*t 0501616

[67] Matthieu Wyart, Leonardo E. Silbert, Sidney R. Nagel, Thomas A. Witten, cond-mat/0508415, submitted to PRE

[68] K.O. Trachenko, M.T. Dove, M.J. Harris and V. Heine, J. Phys: Cond. Matter **12** 8041-8064 (2000)

[69] K.D. Hammonds, A. Bosenick, M.D. Dove, and V. Heine, Amer. Mineralogist, **83**, 476-479 (1998)

[70] M.T. Dove, V. Heine and K.D. Hammonds, Mineralogical Magazine, **59**, 629-639

[71] K. Trachenko, M. Dove, V. Brazhkin and F.S. El'kin Phys. Rev. Lett., **93**, 135502 (2004)

[72] M.D.Newton, M.O'Keeffe and G.V.Gibbs, Physics and Chemistry of Minerals, **6** 305-312 (1980)

[73] B.B. Karki,, M.C. Warren, L. Stixrude, G.J. Ackland and J. Crain, Phys. Rev. B, **55** 3465 (1997)

[74] D.A. Head, Phys. rev. E, **72**, 021303 (2005)

[75] R.B. Griffiths, Phys. Rev. Lett. **23**, 1719 (1969)

[76] C.A. Angell, Y.Yue, L-M Wang, J.R.D. Copley, S. Borick and S. Mossa, *J.Phys.:Condens. Matter* **15** S 1051 (2003)

[77] E. Duval, L. Saviot, L. David, S. Etienne and J. F. Jal, *Europhys. Lett.*, **63**, 778 (2003)

[78] C.S. O'Hern, S.A. Langer, A.J. Liu, and S.R. Nagel., Phys. Rev. Lett. **86**, 111, 2001.

[79] A. Donev, S. Torquato, and F.H. Stillinger, Phys. Rev. E, **71**, 011105 (2005)

[80] G.Parisi, F.Zamponi, cond-mat/0506445

[81] Hernn A. Makse, Nicolas Gland, David L. Johnson, and Lawrence Schwartz, Phys.Rev. E.**70** 061302 (2004)

[82] C.Maloney and A. Lemaitre, Phys. Rev. Lett. 93, 195501 (2004)

[83] B.A. DiDonna, T.C. Lubensky, cond-mat/0506456

[84] JH Snoeijer, JM van Leeuwen, Phys Rev E, **65**, 051306.

[85] J.H. Snoeijer, T.J.H Vlugt, M. Van Hecke and W. van Saarloos, *Phys. Rev. Lett.* **92**, 054302 (2004)

[86] J.H. Snoeijer, *private communication*

[87] J. D. Goddard, Proc. R. Soc. London, Ser. A **430**, 105 (1990).

[88] H.P. Zhang, H.A. Makse, cond-mat 05013370

[89] R. D. Mindlin, J. App. Math. (ASME) **71**, 259, (1949)

[90] E. Somfai, M. van Hecke, W. G. Ellenbroek, W. van Saarloos, cond-mat 0510506

[91] Tuck C. Choy, *Effective Medium Theory, Principles and Applications*, Oxford University press, (1999)

[92] D.A. Keen and M.T. Dove, J.Phys. : cond. matter **11** 9263 (1999)

[93] Hehlen B, Courtens E, Yamanaka A, Inoue K, Jour of non crystalline solids **307** 87-91 SEP 2002

[94] K. Vollmayr, W. Kob, K. Binder, Phys. Rev. B **54** 15808 (1996)

[95] M.T. Dove, M.J. Harris, A.C. Hannon, J.M. Parker, I.P. Swainson and M. Gambhir, Phys. Rev. Lett. (1997) **78**, 1070-1073.

[96] W.A. Kamitakahara, R.L. Cappelletti, P. Boochland, and F. Gompf, Bull. Am. Phys. Soc. **32**, 812 (1987)

[97] Y. Cai and M.F. Thorpe, Phys. Rev. B.,**40**, 15535 (1989)

[98] J-L Barrat, W. Gotze and A Latz, J. Phys.: Cond. Matter **1** 7163-7170 (1989)

[99] Gotze W. and Sjorgen L., Rep. Prog. Phys.,**241** (1992)

[100] A. Ferguson, B. Fisher, B. Chakraborty, Europhys. Lett., **66**, 277 (2004)

[101] A. Donev, S. Torquato, F.H. Stillinger, and R. Connelly, J. Compt. Phys., **197**, 139 (2004)

[102] W. Kob W, JL. Barrat, F. Sciortino,. P. Tartaglia J., Phys. Condensed Matter **12** 6385 (2000)

[103] A.Duri, P Ballesta, L. Cipelletti, H. Bissig and V. Trappe, Fluctuation and Noise Lett.,**5**, 1-15, (2005)

[104] L Buisson, L Bellon and S Ciliberto, J. Phys.: Condens. Matter **15** S1163S1179 (2003)

[105] E. Somfai, J-N Roux, J. Snoejer, M. Van Hecke and W. Van Saarloos, cond-mat 0408128

[106] Ch Liu, S.R. Nagel, Jour. of phys.-cond. matter, **6**, 433-436 (1994)

[107] M. Foret, E. Courtens, R. Vacher, and J-B Suck, Phys. Rev. Lett. **77**, 3835 (1996)

[108] Daniel J. Lacks, Phys. Rev. Lett., **84**, 4629, (2000)

[109] J Kurchan and L. Laloux, J. Phys. A: Math Gen. **29** 1929 (1996)

[110] T.R. Kirkpatrick, D. Thirumalai and P.G. Wolynes, Phys. Rev. A, **40**, 1045, (1989)

[111] L.F. Cugliandolo, J. Kurchan, Phys. Rev. Lett., **71** 173 (1993)

[112] L.V. Woodcock, C.A. Angell, and P. Cheeseman, J. Chem. Phy.s **65**, 1565 (1976); S.A. Brawer, Phys. Rev. Lett. **46**, 778 (1981); J. Chem. Phys. **75**, 3516 (1981); D.A Litton and S.H Garofalini, J. Non-Cryst. Solids **217**, 250 (1997).

[113] Yves Guissani and Bertrand Guillot, Journal of Chemical Physics **104** 7633-7644 (1996)

[114] A. Cavagna, Euro. Phys. Lett. **53**, 490

[115] J.-L. Barrat, J. Badro, and Ph. Gillet, J. Comp. Simul. 20:17–25 (1997).

[116] L. Berthier, Phys. Rev. E, **69**, 020201

[117] B. Bernu et al., Phys. Rev. A 36, 4891 (1987);

[118] E. Kolb, J. Cviklinski, J. Lanuza, P. Claudin, and E. Clement Phys. Rev. E **69**, 031306 (2004)

[119] Jean-Philippe Bouchaud and Giulio Biroli. J. Chem. Phys. **121**, 7347 (2004)

Dynamical susceptibility of glass formers: Contrasting the predictions of theoretical scenarios

Cristina Toninelli,[1] Matthieu Wyart,[2] Ludovic Berthier,[3] Giulio Biroli,[4] and Jean-Philippe Bouchaud[2,5]
[1]*ENS 24 rue Lhomond, 75231 Paris Cedex 05, France*
[2]*Service de Physique de l'État Condensé Orme des Merisiers—CEA Saclay, 91191 Gif sur Yvette Cedex, France.*
[3]*Laboratoire des Verres UMR 5587, Université Montpellier II and CNRS, 34095 Montpellier, France*
[4]*Service de Physique Théorique Orme des Merisiers—CEA Saclay, 91191 Gif sur Yvette Cedex, France*
[5]*Science & Finance, Capital Fund Management 6-8 Bd Haussmann, 75009 Paris, France*
(Received 15 December 2004)

We compute analytically and numerically the four-point correlation function that characterizes nontrivial cooperative dynamics in glassy systems within several models of glasses: elastoplastic deformations, mode-coupling theory (MCT), collectively rearranging regions (CRR's), diffusing defects, and kinetically constrained models (KCM's). Some features of the four-point susceptibility $\chi_4(t)$ are expected to be universal: at short times we expect a power-law increase in time as t^4 due to ballistic motion (t^2 if the dynamics is Brownian) followed by an elastic regime (most relevant deep in the glass phase) characterized by a t or $\sqrt{t}$ growth, depending on whether phonons are propagative or diffusive. We find in both the β and early α regime that $\chi_4 \sim t^{\mu}$, where μ is directly related to the mechanism responsible for relaxation. This regime ends when a maximum of χ_4 is reached at a time $t=t^*$ of the order of the relaxation time of the system. This maximum is followed by a fast decay to zero at large times. The height of the maximum also follows a power law $\chi_4(t^*) \sim t^{*\lambda}$. The value of the exponents μ and λ allows one to distinguish between different mechanisms. For example, freely diffusing defects in $d=3$ lead to $\mu=2$ and $\lambda=1$, whereas the CRR scenario rather predicts either $\mu=1$ or a logarithmic behavior depending on the nature of the nucleation events and a logarithmic behavior of $\chi_4(t^*)$. MCT leads to $\mu=b$ and $\lambda=1/\gamma$, where b and γ are the standard MCT exponents. We compare our theoretical results with numerical simulations on a Lennard-Jones and a soft-sphere system. Within the limited time scales accessible to numerical simulations, we find that the exponent μ is rather small, $\mu<1$, with a value in reasonable agreement with the MCT predictions, but not with the prediction of simple diffusive defect models, KCM's with noncooperative defects, and CRR's. Experimental and numerical determination of $\chi_4(t)$ for longer time scales and lower temperatures would yield highly valuable information on the glass formation mechanism.

DOI: XXXX PACS number(s): 64.70.Pf

I. INTRODUCTION

The idea that the sharp slowing down of supercooled liquids is related to the growth of a cooperative length scale dates back at least to Adam and Gibbs [1]. But it is only a few years back that this idea has started being substantiated by convincing experiments [2–6], numerical simulations [7–14], and simple microscopic models [15–25,27]. One of the basic problems has been to find an observable that allows one to define and measure objectively such a cooperative length scale. An interesting quantity, proposed a few years ago in the context of mean-field p-spin glasses [28] (see [29] for an important early insight) and measured in simulations, is a four-point density correlator, defined as

$$G_4(\vec{r},t) = \langle \rho(0,0)\rho(0,t)\rho(\vec{r},0)\rho(\vec{r},t)\rangle - \langle \rho(0,0)\rho(0,t)\rangle\langle \rho(\vec{r},0)\rho(\vec{r},t)\rangle, \quad (1)$$

where $\rho(\vec{r},t)$ represents the density fluctuations at position $\vec{r}$ and time t. In practice one has to introduce an overlap function w [28] to avoid a singularity due to the evaluation of the density at the same point or consider slightly different correlation functions [30]. This quantity measures the correlation in space of local-time correlation functions. Intuitively, if at point 0 an event has occurred that leads to a decorrelation of the local density over the time scale t, $G_4(\vec{r},t)$ measures the probability that a similar event has occurred a distance $\vec{r}$ away within the same time interval t (see, e.g., [31]). Therefore $G_4(\vec{r},t)$ is a candidate to measure the heterogeneity and cooperativity of the dynamics. The best theoretical justification for studying this quantity is to realize that the order parameter for the glass transition is already a two-body object—namely, the density-density correlation function $C(t)=\langle \rho(0,0)\rho(0,t)\rangle$—which decays to zero in the liquid phase and to a constant value in the frozen phase. The four-point correlation $G_4(\vec{r},t)$ therefore plays the same role as the standard two-point correlation function for a one-body order parameter in usual phase transitions. Correspondingly, the associated susceptibility $\chi_4(t)$ is defined as the volume integral of $G_4(\vec{r},t)$ and is equal to the variance of the correlation function [28,32,33]. The susceptibility $\chi_4(t)$ has been computed numerically for different model glass formers and indeed exhibits a maximum for $t=t^* \sim \tau_\alpha$, the relaxation time of the system [11–14]. The peak value $\chi_4(t^*)$ is seen to increase as the temperature decreases, indicating that the range of $G_4(\vec{r},t^*)$ increases as the system becomes more sluggish. The dynamical correlation length $\xi_4(t^*)$ extracted from $G_4(\vec{r},t^*)$ in molecular dynamics simulations grows and becomes of the order of roughly 10 interparticle distances when the time scale is of the order of 10^5 microscopic time scales τ_0 with $\tau_0 \sim 0.1$ ps for an atomic liquid. In experiments close

to the glass transition the dynamical correlation length has been found to be only slightly larger, between 10 and 20 interparticle distances [2,4]. This is puzzling because experiments are done on systems with relaxation times that are several orders of magnitude larger than in simulations. In fact, extrapolating simulation results in the experimental regime would lead to much larger dynamical correlation lengths. The origin of this puzzle is still unclear; see Ref. [18] for a recent discussion. Experiments on dynamical heterogeneity bridging the gap between numerical and macroscopic time scales would be extremely valuable to resolve this paradox.

Several scenarios have been proposed to understand the existence of nontrivial dynamical correlations and their relation to thermodynamical singularities. Adam and Gibbs [1], Kirkpatrick *et al.* [34] (for a different formulation, see Ref. [35]), and Kivelson and Tarjus [36] have proposed, using somewhat different arguments, the idea of collectively rearranging regions (CRR's), of size ξ, that increase as the temperature is decreased. The evolution of the system is such that these regions are either frozen or allowed to temporarily and collectively unjam for a short time until a new jammed configuration is found.

In apparent contradiction with the existence of the growing length scale, the mode-coupling theory (MCT) of glasses states that the self-consistent freezing of particles in their cages is a purely local process with no diverging length scale at the transition [37]. However, this point of view is in disagreement with the results found for mean-field disordered systems [28,29] that are conjectured to provide a mean-field description of the glass transition and display an MCT-like dynamical transition. Indeed it was recently shown that within MCT $G_4(\vec{r},t)$ in fact develops long-range correlations close to the critical MCT temperature T_c [32]. Within a phase-space interpretation of the MCT transition, the mechanism for this cooperative behavior for $T>T_c$ is the progressive rarefaction of energy lowering directions [38]. Within a real-space interpretation, the MCT transition is due to the formation of a large number of metastable states, each one characterized by a surface tension that increases from zero at T_c. As one approaches T_c from above, the relevant eigenvectors of the dynamical Hessian become more and more extended, which means that the modes of motion that allow the system to decorrelate are made of very-well-defined, collective rearrangements of larger and larger clusters of particles (see the recent work of Montanari and Semerjian [39]). For smaller temperatures $T<T_c$, "activated events" are expected to play a crucial role. They are believed to be responsible for the destruction of the freezing transition at T_c. This regime has been tentatively described by adding "hopping terms" in the MCT equations [37] or within a CRR scenario [34,35]. Note that the random first-order theory of [34] unifies MCT with CRR's predicting a first temperature regime (close to T_{MCT}) where MCT applies and then a crossover toward CRR's (the mosaic state) that describe the physical behavior close to the Kauzman temperature.

Exploiting yet a different set of ideas, models of dynamical facilitation, such as the FredericksonAndersen [19] or Kob-Andersen models [24], have recently been proposed as paradigms for glassy dynamics [15,20,23]. In these models, the motion of particles is triggered by "mobility defects" that diffuse and possibly interact within the system. As the temperature is lowered or the density is increased, the concentration of defects goes down and the relaxation time of the system increases. The dynamics is obviously heterogeneous since it is catalyzed by defects that cannot be everywhere simultaneously. The characteristic length scale in this case is related to the average distance between defects to some model- and dimension-dependent exponent [15,18,20,23,25]. The ideas behind these models are somehow similar to the one of free-volume theories and can be traced back to the first explanation of slow dynamics in terms of defects motion [26]. Kinetically constrained models have the important merit of showing how from simple local microscopic rules a relaxation governed by the diffusion (or subdiffusion) of nontrivial defects may arise.

Understanding the mechanism behind the growth of the dynamical correlation length is certainly an important step—arguably the most important one—to understand the cause of the slowing down of the dynamics. Furthermore, the different scenarios for the glass transition can be tested, contrasting their quantitative prediction for the four-point correlation function $G_4(\vec{r},t)$ to the numerical, and hopefully soon experimental, results. Following these premises we investigate in this paper the analytical shape of $G_4(\vec{r},t)$ for several simple models. We show that $G_4(\vec{r},t)$ indeed contains some important information concerning the basic relaxation mechanisms. However, we show that, perhaps disappointingly, models where cooperativity is absent or trivial lead to four-point correlation functions and dynamical susceptibilities χ_4 that exhibit nontrivial features. Other, more complex observables will have to be defined to really grasp the nature of the collective motions involved in the relaxation process of glasses [8,40].

Let us summarize the main results of our study in terms of the susceptibility $\chi_4(t)$ and time sectors. In a supercooled liquid there are separate regimes of time scales corresponding to different physical behavior (see Fig. 1). On microscopic time scales particles move ballistically if the dynamics is Newtonian or diffusively if the dynamics is Brownian. On a longer time scale, interactions start playing a role, which can be described approximately using elasticity theory, before a truly collective phenomenon sets in. This nontrivial glassy regime is the β regime, within which correlation functions, such as, for example, the dynamical structure factor, develop a plateau. The β regime is divided further in an early- and a late-β regime corresponding, respectively, to the approach and departure from the plateau of the correlation function. Finally the structural relaxation time scale on which correlation functions decay to zero is the α regime. All previous studies have focused on the behavior of $\chi_4(t)$ at times of the order of τ_α which correspond to the peak of $\chi_4(t)$. We show that $\chi_4(t)$ has in fact a rich structure in time and different behavior in different time sectors. In many of these regimes, $\chi_4(t)$ behaves as a power law of time t^μ with different values of μ. During the ballistic time scale one finds $\mu=4$ ($\mu=2$ for Brownian dynamics), whereas during the elastic regime (most relevant deep in the glass phase), the exponent becomes $\mu=1$ for ballistic phonons and $\mu=1/2$ for

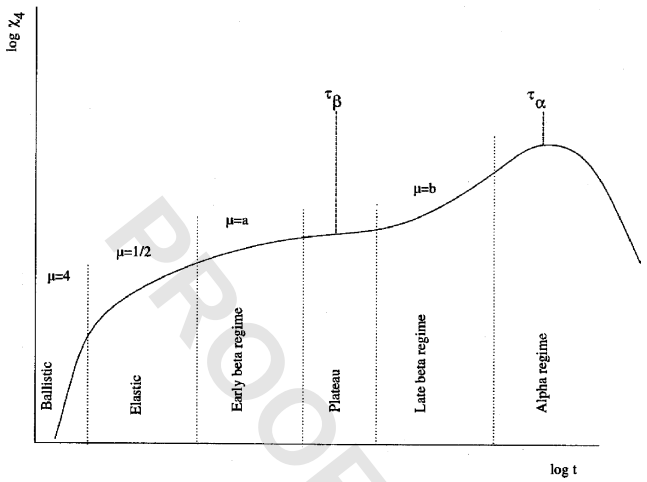

FIG. 1. Sketch of the time behavior of $\chi_4(t)$, with all the different time regimes, within the MCT description that we find to be a good description around T_c. As the temperature is lowered, we expect the elastic regime to extend up to τ_α.

diffusive phonons. The behavior in the β and α regimes is intimately related to the physical mechanism for relaxation and indeed we find quite different answers depending on which scenario we focus on. MCT predicts exponents $\mu=a$ and $\mu=b$ on time scales corresponding, respectively, to the early- and late-β regimes, where a and b are the standard MCT exponents obtained from the study of the dynamical structure factor. The power law t^b extends until the peak in $\chi_4(t)$ is reached.

The other scenarios only make predictions in the α regime. In the case of CRR's one has $\chi_4 \sim t$ or $\chi_4 \sim (\ln t)^{d+1/\psi}$ before the peak depending whether one assumes that the relaxation occurs via bulk nucleation events or domain wall fluctuations; see below. For diffusing defects in dimension $d=3$, the exponent is $\mu=2$. If defects have a nontrivial diffusion exponent z, such that their displacement at time t scales as $t^{1/z}$, then $\mu=2d/z$ for $d<z$ and $\mu=2$ otherwise. The overall behavior of $\chi_4(t)$ is summarized by Fig. 1, which specializes to the MCT predictions for simplicity.

Another important feature of χ_4 is the growth of the peak compared to the growth of the time $t=t^* \sim \tau_\alpha$ at which the peak takes places [21]. This is found to scale as $\chi_4(t^*) \sim t^{*\lambda}$, with $\lambda=0$ (logarithm) for CRR's, $\lambda=1$ for freely diffusing defects, $\lambda=d/z$ for anomalously diffusing defects for $d<z$, and $\lambda=1$ again for $d>z$. Note that if the defect diffusion coefficient itself scales with t^* as $1/t^{*f}$, such as, for example, in the one-spin facilitated FA model, there is an extra contribution that gives $\lambda=1-f$ for $d>z$. Finally, one has $\lambda=1/\gamma$ in the context of MCT, where γ describes the power-law divergence of the relaxation time as the critical MCT temperature is approached.

We have checked these predictions in two model systems of glass-forming liquids: a Lennard-Jones and a soft-sphere mixture. Concerning the behavior of $\chi_4(t)$ in the late-β and -α regimes, the most interesting time sectors, we have found reasonable agreement with the MCT predictions for four point correlators. This agreement is by no means trivial and is actually quite unexpected unless MCT indeed captures some of the physics of the problem. Instead models of diffusing, defects do not describe well the numerical results. This is perhaps not very surprising since we are focusing on two fragile liquids (at least in the numerical time window) at temperatures well above the experimental glass transition. It might be that the predictions of these models work only on larger time scales. In any case, we expect instead that for strong liquids displaying an Arrhenius behavior the predictions for $\chi_4(t)$ obtained studying the model of simple diffusing defects might hold quantitatively, since it seems quite well established from numerical simulations that relaxation in strong liquids is triggered by the diffusion of connectivity defects [41,42]. Finally, the CRR picture does not agree quantitatively with our present numerical data. However, this picture is supposed to describe the liquid dynamics precisely in the low-temperature and long-time regime, which is presently beyond numerical capabilities. Again, experimental results probing the behavior of $\chi_4(t)$ in this regime would be highly valuable to put strong constraints on the different theoretical scenarios of glass formation.

The organization of the paper is as follows. In Sec. II we discuss the behavior of $\chi_4(t)$ on microscopic time scales. Then, we analyze the predictions of elasticity theory in Sec. III. In Secs. IV and V we focus on the behavior of $\chi_4(t)$ in the β and α regimes for MCT and CRR's. In Sec. VI we discuss the predictions of defect models analytically using an independent defect approximation and by numerical simulations of kinetically constrained models. In Sec. VII we compare the different predictions to the results of numerical simulations of models of glass-forming liquids. We present our conclusions in Sec. VIII.

II. MICROSCOPIC DYNAMICS

On very short time scales the behavior of χ_4 can be computed exactly. For simplicity, we characterize the dynamics through the self-intermediate scattering function

$$F_s(k,t) = \frac{1}{N}\sum_i \langle \cos \vec{k} \cdot [\vec{r}_i(t) - \vec{r}_i(0)] \rangle \tag{2}$$

and define the dynamic susceptibility as the variance of the fluctuations of $F_s(k,t)$:

$$\chi_4(t) = N\left[\left\langle \left(\frac{1}{N}\sum_i \cos \vec{k} \cdot [\vec{r}_i(t) - \vec{r}_i(0)] \right)^2 \right\rangle - \left\langle \frac{1}{N}\sum_i \cos \vec{k} \cdot [\vec{r}_i(t) - \vec{r}_i(0)] \right\rangle^2 \right]. \tag{3}$$

The full intermediate four-point scattering function defined in Eq. (1) in fact contains very similar information, even for interacting systems—as shown by numerical simulations [12,13].

On a very short time scale particles move ballistically if the dynamics is Newtonian, $\vec{r}_i(t) - \vec{r}_i(0) = \vec{v}_i t + O(t^2)$, where $\vec{v}_i$ is the velocity of the particle i at time t. Since the system is in equilibrium all the $\vec{v}_i$'s are independent Gaussian variables with variance $\langle \vec{v}_i \cdot \vec{v}_j \rangle = \delta_{ij} 3k_B T/m$, where T is the tempera-

ture, m the mass of the particles, and k_B the Boltzmann constant. Using this property it is straightforward to obtain

$$F_s(k,t) = \exp\left(-\vec{k}^2 \frac{k_B T}{2m} t^2\right) \tag{4}$$

and

$$\chi_4(t) = F_s(k,t)^2 \left[\cosh\left(-2\vec{k}^2 \frac{k_B T}{m} t^2\right) - 1\right]. \tag{5}$$

For an interacting particle systems this is only valid on short time scales—for example, smaller than the collision time for short-ranged interactions. This leads to an initial power-law increase that reads

$$\chi_4(t) = \frac{1}{2}(\vec{k}^2)^2 \left(\frac{k_B T}{m}\right)^2 t^4 + O(t^6). \tag{6}$$

Note that if one had chosen Langevin dynamics (i.e., $\partial_t \vec{r}_i = \partial_{\vec{r}} H + \vec{\eta}_i$) instead of Newtonian dynamics, Eqs. (5) and (6) would have been identical except for the replacement of $k_B T t^2/m$ by $2Tt$, again for small times. Thus changing from Newtonian to Langevin dynamics, the initial power-law increase of $\chi_4(t)$ changes from t^4 to t^2. This is similar to the change in the mean-square displacement that increases as t^2 and t, respectively, for Newtonian and Langevin dynamics.

In the above example, however, it is clear that the increase of χ_4 with time has nothing to do with the increase of a correlation length, since particles are assumed to be independent. In other words, the four-point correlation $G_4(\vec{r},t)$ has a trivial δ-function spatial dependence, but the height of the δ peak increases with time. As will be discussed later in the paper, it is important to normalize $\chi_4(t)$ by the value of $G_4(\vec{r}=0,t)$ to conclude from the four-point susceptibility that a length scale is indeed growing in the system.

III. ELASTIC CONTRIBUTION

For longer time scales the interaction between particles starts playing a role. Generically one expects that in the time regime where the displacements of particles remain small, an elastic description should be valid. In a solid or in a glass deep below T_g, there is no further relaxation channels and the elastic contribution to χ_4 should be the only relevant one. In a supercooled liquid around the mode-coupling temperature T_c, the elastic regime is interrupted by the collective β regime, where in some sense phonon-phonon interactions completely change the physical picture. Although we expect such a crossover, we have at present no detailed theoretical description of it.

In the following we analyze again the behavior of the four-point self-intermediate scattering function assuming that the dynamical behavior of the liquid can be described, within a restricted time sector, as an elastic network (we will discuss later how to include, in a phenomenological way, viscous flow). Perhaps surprisingly, we find a nontrivial structure for G_4 in this model, with an ever growing "cooperative" length scale which comes from the dynamics of phonons, which represents the simplest form of cooperativity.

We consider an isotropic solid immersed in a viscous thermal bath. The energy of the system is given by

$$H = \int d^d r \frac{1}{2} \kappa_1 \left[\sum_i u_{ii}\right]^2 + \kappa_2 \sum_{i,j} u_{i,j}^2, \tag{7}$$

where κ_1, κ_2 are the Lamé coefficient, $u_{i,j} = \frac{1}{2}[d\phi_i/dx_j + d\phi_j/dx_i]$ is the deformation tensor, and $\vec{\phi}$ the displacement field from an undeformed reference state. Note that $\vec{\phi}(x)$ is simply the continuum limit of the displacement of each particle with respect to its equilibrium (bottom of the well) position.

As is well known, the above energy leads to three independent phonon modes (one longitudinal and two transverse modes). For simplicity, we only consider one deformation mode and write the Hamiltonian in Fourier space as

$$H = \frac{1}{2} \kappa \int \frac{d^d k}{(2\pi)^d} k^2 \phi_k \phi_{-k}, \tag{8}$$

where κ is an effective elasticity modulus. The mode k has an energy $E_k = \kappa k^2 \phi_k \phi_{-k}/2$ and therefore we expect, in equilibrium, $\langle \phi_k \phi_{-k} \rangle = T/\kappa k^2$, where the Boltzmann constant has been set to unity. Our goal is to calculate the dynamical correlation functions of the system. We describe the dynamics by a Langevin equation with a local noise:

$$m \frac{\partial^2 \phi(\vec{r},t)}{\partial t^2} + \nu \frac{\partial \phi(\vec{r},t)}{\partial t} = \kappa \Delta \phi(\vec{r},t) + \zeta(\vec{r},t), \tag{9}$$

where $\zeta(x,t)$ is a Gaussian noise uncorrelated in space and time, of variance equal to $2\nu T$. Taking the Fourier transform

$$m \frac{\partial^2 \phi_k}{\partial t^2} + \nu \frac{\partial \phi_k}{\partial t} = -\kappa k^2 \phi_k + \zeta_k(t), \tag{10}$$

$\zeta_k(t)$ is again a Gaussian noise uncorrelated for different k's and time.

In this section, we only consider in details the overdamped case $m=0$ and set $D = \kappa/\nu$, but also give at the end the result for the purely propagative case $\nu=0$ (see also Appendix A). One easily deduces the non-equal-time correlation in the overdamped case:

$$\langle \phi_k(t) \phi_{-k}(0) \rangle = \frac{T}{\kappa k^2} e^{-Dk^2 t}. \tag{11}$$

Let us now define the function

$$F^{(q)}(r,t) = \sum_i \delta(r - r_i(0)) \cos\{q[r_i(t) - r_i(0)]\}, \tag{12}$$

whose average equals the self-intermediate scattering function up to a constant (the particle density).

Using the microscopic definition of $\vec{\phi}$ we obtain that

$$C(q,t) = \langle F^{(q)}(r,t) \rangle \simeq \langle e^{iq[\phi(\vec{r},t) - \phi(\vec{r},0)]} \rangle = e^{-q^2 \langle [(\phi(\vec{r},t) - \phi(\vec{r},0)\ ]^2 \rangle/2}, \tag{13}$$

where the last equality comes from the Gaussian nature of the deformation field. Using the above results on the correlation of the Fourier modes, we find

$$\langle [\phi(\vec{r},t) - \phi(\vec{r},0)]^2 \rangle = \frac{2T}{\kappa} \int \frac{1 - e^{-Dk^2 t}}{k^2} \frac{d^d k}{(2\pi)^d}. \quad (14)$$

As is well known, this integral behaves differently for $d \leqslant 2$ and for $d > 2$, reflecting the fact that phonons destroy translational order in low dimensions. As above, we will only consider here the physical case $d=3$, relegating the discussion of the other cases to Appendix A. For $d=3$, we need to introduce an ultraviolet cutoff Λ on the wave vector k, which is the inverse of the underlying lattice spacing a. Then, the above integral goes to a constant $\propto \Lambda$ at large times, reflecting the fact that particles are localized in their "cage." Therefore, the self-intermediate scattering function $C(q,t)$ decays at small times $\Lambda^2 Dt \ll 1$ before saturating to a "plateau" value given by

$$f_q \equiv C(q, t \to \infty) = \exp\left(- c \frac{T \Lambda q^2}{\kappa} \right), \quad (15)$$

where c is a numerical constant. (Note that $T\Lambda q^2/\kappa$ has no dimension and is expected, from a Lindemann criterion, to be of the order of 0.05 at half the melting temperature and for $q=\Lambda$.) In real glass-forming liquids, this plateau phase does not persist forever, and $C(q,t)$ finally decays to zero beyond $t=\tau_\alpha$, in the so-called α-relaxation regime. A modification of the model to account for this decorrelation will be discussed later. Furthermore, the above pseudo-β regime predicted by elastic theory does not explain quantitatively the β regime in supercooled fragile liquids, except probably on relatively short time scales—say, up to a few picoseconds. On the other hand, at temperatures below T_g or for strong glasses, we expect that the elastic regime will extend up to τ_α and compete with other mechanisms, such as the defect-mediated correlation discussed in Sec. VI below.

The calculation of $G_4^{(q)}(\vec{r},t) = \langle F^{(q)}(r',t) F^{(q)}(r'+r,t) \rangle_c$ is detailed in Appendix A. One immediately sees that $G_4^{(q)}(\vec{r},t)$ is governed by a diffusive correlation length $\xi(t) \sim \sqrt{Dt}$ with $D = \kappa/\nu$, as expected from the structure of the Langevin equation that describes relaxational dynamics. Clearly, in the case of propagative phonons, one finds $\xi(t) \sim Vt$ with $V^2 = \kappa/m$. The final result, see Appendix A, is

$$G_4^{(q)}(\vec{r},t) = C^2(q,t)\{\cosh[2q^2 R(\vec{r},t)] - 1\}, \quad (16)$$

where

$$R(\vec{r},t) = \frac{T}{\kappa} (Dt)^{1-d/2} F\left(\frac{r}{\sqrt{Dt}} \right) \quad (17)$$

and we find (see Appendix A) $F(z) \simeq (4\pi z)^{-1}$ for $z \ll 1$ and $F(z) \simeq (2\pi^{3/2})^{-1} \exp(-z^2/4)/z^2$ for $z \gg 1$. Note the similarity between the expression in Eq. (16) and the corresponding one (5) derived in the previous section. One can check that indeed the short-time behavior is indeed the one derived before in the case of Langevin dynamics for the particles, as expected. Let us now focus on long times, but still within the elastic regime, $\Lambda^2 Dt \gg 1$, and for $r \ll \xi(t)$,

$$G_4^{(q)}(\vec{r},t) = f_q^2 \left[\cosh\left(\frac{Tq^2}{2\pi\kappa r} \right) - 1 \right]. \quad (18)$$

Suppose for simplicity that we are in a regime where the argument of the cosh is always small, corresponding to the limit $Tq^2\Lambda \ll \kappa$ (remember that by definition $r > a = 2\pi/\Lambda$, where a is the interatomic distance). Then, $G_4(\vec{r},t) \sim r^{-2}$ for $\Lambda^{-1} \ll r \ll \xi(t)$. For larger scales $r \gg \xi(t)$ decays as a Gaussian—i.e., superexponentially fast. Note that the small-r behavior of $G_4(\vec{r},t)$ is not of the Ornstein-Zernike form ($1/r$ in $d=3$). Integrating G_4 over $\vec{r}$ we find the dynamical susceptibility

$$\chi_4^{(q)}(t) \sim \frac{T^2 q^4 f_q^2}{\kappa^2} \xi(t). \quad (19)$$

This result is actually valid both for in the diffusive limit where $\xi(t) = \sqrt{Dt}$ and in the propagative regime where $\xi(t) = Vt$. Therefore $\chi_4^{(q)}(t)$ increases either as $\sqrt{t}$ or as t (note that in the limit of small times one recovers the t^4 or t^2 laws obtained in the previous section). In the general case, one expects a crossover between a propagative regime at small times $t < m/\nu = D/V^2$ (of the order of ps in glass formers; see [43]) and a diffusive regime for longer time scales. Thus, looking at $\chi_4^{(q)}(t)$ as a function of time in a log-log plot one should see first a straight line corresponding to the ballistic or diffusive motion leading, respectively, to slope $\mu=4$ or $\mu=2$, bending over toward a smaller slope (1 or 1/2, or both, depending on the strength of the dissipation). The order of magnitude of $\chi_4^{(q)}(t)$, as given by Eq. (19), can be estimated to be $\sim(10^{-3}-10^{-2}) a^2 \xi(t)$ for $q=\Lambda$. In the propagative regime with $t=1$ ps, $V = 3 \times 10^3$ m/s, and $a=0.3$ nm, one finds $\xi = 10a$ and $\chi_4^{(q)} \sim (10^{-2}-10^{-1}) a^3$—i.e., a small, but perhaps detectable signal from the phonons. Only on much larger time scales will the elastic contribution be significant, a regime that can be reached deep in the glass phase [44]. As mentioned above, other collective modes come into play in supercooled fragile liquids, in particular around the mode-coupling temperature, and give rise to the β regime where "cages" themselves become more complex, extended objects [32].

The above calculation shows that in an elastic solid with diffusive or propagative phonon modes, the dynamical susceptibility increases without bound, reflecting the presence of Goldstone soft modes in the system. Of course, in a real glass the correlation function $C^{(q)}(t)$ eventually decays to zero beyond the α-relaxation time τ_α, as particles start diffusing out of their cages, far away from their initial position. If phonons were the only relevant excitations, this would cause the dynamical susceptibility to peak around $t = t^* = \tau_\alpha$. A phenomenological model that describes the decay of $\chi_4^{(q)}(t)$ within the above elastic framework is to assume a (Maxwell) viscoelastic local modulus:

$$\frac{\partial \phi(\vec{r},t)}{\partial t} = \kappa \left[\int_{-\infty}^{t} dt' e^{-\gamma(t-t')} \Delta \frac{\partial \phi(\vec{r},t')}{\partial t'} \right] + \zeta(\vec{r},t), \quad (20)$$

with $\gamma \sim \tau_\alpha^{-1}$, corresponding to a frequency-dependent elastic modulus $\kappa(\omega) = i\kappa\omega/(i\omega + \gamma)$. In this model, the dynamics of

ϕ becomes diffusive at times $>\gamma^{-1}$ and the dynamic structure factor therefore decays exponentially beyond that time. Of course, the model itself becomes inconsistent at large times, since the underlying lattice needed to define the deformation field ϕ has by then totally melted.

The conclusion of this section, however, is that since supercooled liquids behave at high frequencies ($\omega \gg \gamma, \tau_\alpha^{-1}$) like solids, the four-point correlation and dynamical susceptibility are expected to reveal, in a certain time domain, a nontrivial behavior unrelated to the structure of the "collective processes" discussed below (MCT, diffusive defects, CRR's) that one usually envisions to explain glassy dynamics.

IV. MODE-COUPLING THEORY

As mentioned in the Introduction the mode-coupling theory of supercooled liquids predicts the growth of a cooperative length as the temperature is decreased or the density increased [28,29,32] and makes detailed predictions on the shape of $\chi_4(t)$. The four-point correlation function becomes critical near the mode-coupling transition temperature T_c. The behavior of the susceptibility $\chi_4(t)$ is encoded in ladder diagrams [29,32]. From the analytical and numerical results of [32] and analyzing the ladder diagrams [32,45], we have found that, in the β regime,

$$\chi_4(t) \sim f_1(t\epsilon^{1/2a})/\sqrt{\epsilon}, \quad t \sim \tau_\beta, \tag{21}$$

and in the α regime,

$$\chi_4(t) \sim f_2(t\epsilon^{1/2a+1/2b})/\epsilon, \quad t \sim \tau_\alpha, \tag{22}$$

where $\epsilon = T - T_c$, a, b, and $\gamma = 1/2a + 1/2b$ are the MCT exponents for the dynamical structure factor, and $f_1(x)$ and $f_2(x)$ are two scaling functions. Requiring that the dependence on ϵ drop out when $t\epsilon^{1/2a} \ll 1$ one finds that $f_1(x) \sim x^a$ when $x \ll 1$. This leads to a power-law behavior $\chi_4 \sim t^a$ in the early-β regime—i.e., when the intermediate scattering functions approaches a plateau. In the same way, matching the behavior of f_1 when $x \gg 1$ to the one of f_2 when $x \ll 1$ one finds another power-law behavior $\chi_4 \sim t^b$ on time scales between the departure from the plateau and the peak of χ_4. We give in Fig. 1 a schematic summary of the shape of $\chi_4(t)$ within the MCT description of supercooled liquids.

Finally, as discussed in [32], at times $t = t^* \sim \tau_\alpha$, χ_4 reaches a maximum of height $(T - T_c)^{-1}$. Using the relation $\tau_\alpha \sim (T - T_c)^{-\gamma}$, valid within MCT, one finally finds $\chi_4(t^*) \sim t^{*1/\gamma}$.

V. COLLECTIVELY REARRANGING REGIONS

Under the term CRR, we gather many similar scenarios that differ in their details, as discussed in the Introduction [1,34–36]. Within the frustration-limited domains scenario of Ref. [36] it seems natural to envision the dynamics as the activated motion of domains pinned by self-generated disorder. In the case of the random first-order theory of Refs. [34,35], the details of the decorrelation mechanism are not entirely clear. There should be, on the one hand, activated fluctuations of domain walls between different states, again pinned by self-generated disorder. However, the fluctuations leading to a change of state may be the nucleation of a completely different state starting from the bulk. The latter process can be modeled as a nearly instantaneous event with a certain (small) nucleation rate. In the following we shall analyze separately these two types of fluctuations and their consequences on the shape of $\chi_4(t)$.

A. Instantaneous events

Suppose that the dynamics is made of nearly instantaneous events that decorrelate the system in a compact "blob" of radius ξ_0. The probability per unit time and volume for such an event to appear around site $\vec{r}$ is Γ. We compute the four-body correlation of the persistence, $n_r(t)$, defined to be equal to one if no event happened at $\vec{r}$ between times 0 and t and equal to zero otherwise. The four-body correlation is then defined as

$$G_4(\vec{r},t) = \langle n_r(t) n_0(t) \rangle - \langle n_r(t) \rangle^2. \tag{23}$$

Clearly, the averaged correlation function $C(t) = \langle n_r(t) \rangle$ is simply given by $C(t) = \exp(-\Omega \Gamma \xi_0^d t)$ where Ω is the volume of the unit sphere. For $G_4(\vec{r},t)$ to be nonzero, an event must have happened simultaneously at $\vec{r}$ and at 0, leading to

$$G_4(\vec{r},t) = C^2(t)\{\exp[\Gamma t \xi_0^d f(r/\xi_0)] - 1\}, \tag{24}$$

where $f(x)$ is the volume of the intersection between two spheres of unit radius with centers at distance x apart. Clearly, $f(x>2) = 0$. Therefore, $G_4(\vec{r},t)$ is nonzero only if $r < 2\xi_0$, and is in fact roughly constant there. For a given r satisfying this bound, G_4 first grows linearly with time, reaches a maximum for $t = t^* \approx \Gamma^{-1} \xi_0^{-d}$ and decays exponentially beyond that time. The same behavior is found for $\chi_4(t)$, which grows initially as t^μ with $\mu = 1$ and reaches a maximum such that $\chi_4(t^*) \propto \xi_0^d$. Assuming finally that these events are activated [34,35], with a barrier growing like $\Upsilon \xi_0^\psi$, where ψ is a certain exponent, one expects $t^* \sim \tau_0 \exp(\Upsilon \xi_0^\psi / T)$, and therefore $\chi_4(t^*) \propto (\ln t^*)^{d/\psi} \propto \xi_0^d$.

The rearranging regions could have of course more complicated shapes than the simple sphere assumed above. As long as these objects are reasonably compact, the above results will still hold qualitatively. On the other hand, if these regions are fractal with a dimension $d_f < d/2$, the above results on G_4 will hold with the argument in the exponential given by $\Gamma t r^{2d_f - d}$; one also finds $t^* \approx 1/\Gamma \xi_0^{d_f}$ and $\chi_4(t^*) \propto \xi_0^{d_f}$.

B. Domain wall fluctuations

In this case the picture that we have in mind is similar to the case of a disordered ferromagnet with pinned domain walls, where the typical time to flip a domain is comparable to the interevent time. In that case, an "event" is in fact the slow fluctuation of domain walls that progressively invade the bulk of the domain [in the follow we neglect the fast equilibrium dynamics taking place inside the domains that determines the evolution of $\chi_4(t)$ on short time scales]. The early-time behavior of $\chi_4(t)$ is given by the square of the number of particles that relax per unit volume thanks to the same domain wall (see [31] for the same situation out of

equilibrium in pure systems).[1] Let again ξ_0 be the typical size of a domain and $\ell(t)$ the length scale over which the domain walls fluctuate during time t. Considering that on the surface of each domain there are order $(\xi_0/\ell)^{d-1}$ subdomains of linear size ℓ and that the number of particles in each of these subdomains is proportional to ℓ^d, we get $\chi_4(t) \propto \xi_0^{-d}(\xi_0/\ell)^{d-1}\ell^{2d} \propto \ell^{d+1}/\xi_0$. We are descarding for simplicity both the possibility of fractal domains and that transverse fluctuations behave differently from longitudinal ones. Assuming thermal activation over pinning energy barriers that grow like $\Upsilon\ell^{\psi}$ [46], we finally get $\chi_4(t) \propto \xi_0^{-1}(\ln t)^{d+1/\psi}$. Therefore, in this case, the exponent μ is formally zero and the growth of $\chi_4(t)$ is only logarithmic. The maximum of χ_4 occurs at time t^* such that $\ell(t^*) \approx \xi_0$, which implies that the maximum of the susceptibility also scales logarithmically with t^*, $\chi_4(t^*) \propto \xi_0^{-1}(\ln t^*)^{d+1/\psi} \propto \xi_0^d$. The same scaling of the maximum of the susceptibility with the typical domain size is obtained in nondisordered coarsening systems [31].

The conclusion of the above analysis is that if the CRR relaxation is due to both instantaneous events and domain wall fluctuations, the latter will dominate the time behavior of χ_4 before the peak as can be readily deduced by comparing their relative contributions to $\chi_4(t)$. If for some reason domain walls are particularly strongly pinned and bulk nucleation becomes dominant, then the exponent $\mu=1$ should be observable. The height of the peak, on the other hand, behaves identically in both models. Thus, as the temperature is reduced, one should see a power-law behavior before the peak with an exponent $0<b<1$ in the MCT regime followed by an effective exponent μ either decreasing toward zero or increasing toward one depending on whether the domain wall contribution dominates or not. However, at lower temperatures, the elastic contribution will also start playing a role, which might completely dominate over the CRR contribution. This suggests that other observables, which quantify more specifically the collective dynamics, should be devised to reveal a CRR dynamics.

VI. DEFECT-MEDIATED MOBILITY

A. Independently diffusing defects

As the simplest realization of the defect-mediated scenario for glassy dynamics advocated in [15,16,19,20,24], we consider a lattice model in which mobility defects, or vacancies, perform independent symmetric random walks. We assume for the moment that these vacancies cannot be created or destroyed spontaneously. We shall compute the same function $G_4(\vec{r},t)$ as in Eq. (23) above, arguing that when such a vacancy crosses site $\vec{r}$, the local configuration is reshuffled and the local correlation drops to zero. Therefore, $n_r(t)$ is equal to one, if no vacancy ever visited site $\vec{r}$ between $t=0$ and t, and zero otherwise. Thus, $\langle n_r(t)\rangle$ represents a density-density dynamical correlation function whereas $\langle n_0(t)n_{\vec{r}}(t)\rangle - \langle n_0(t)\rangle^2$ corresponds to $G_4(\vec{r},t)$.

[1]We are implicitly assuming that the variance of N_α, the number of particles that relax per unit volume thanks to the same domain wall, equals the square of the average N_α

From now on we will denote by N_v the number of vacancies, by V the total volume, by $\rho_v = N_v/V = 1-\rho$ the vacancy density and by $P_{\vec{x}}^z(t)$ the probability that a vacancy starts in z at time zero and never reaches x until time t. The probability that a vacancy starts in z at time zero and reaches for the first time x at a time $u \leqslant t$ is therefore $P_x^z(t) = 1 - P_{\vec{x}}^z(t)$.

The computation of $\langle n_x(t)\rangle$ is identical to the target annihilation problem considered in [47]. Since we assume defects to be independent, the defect distribution is uniform and we have

$$\begin{aligned}\langle n_x(t)\rangle &= \left[\frac{1}{V}\sum_{z,z\neq x} P_{\vec{x}}^z(t)\right]^{N_v} \\ &= \left[\frac{1}{V}\sum_{z,z\neq x}(1-P_x^z(t))\right]^{N_v} \\ &= \exp\left[-\rho_v - \rho_v \sum_{z,z\neq x} P_x^z(t)\right]. \end{aligned} \tag{25}$$

The correlation function $\langle n_x(t)n_y(t)\rangle$ can be also expressed in terms of probability distributions of a single random walk in a similar way:

$$\begin{aligned}\langle n_x(t)n_y(t)\rangle &= \left[\frac{1}{V}\sum_{z,z\neq x,y} P_{\vec{x},\vec{y}}^z(t)\right]^{N_v} \\ &= \left[\frac{1}{V}\sum_{z,z\neq x,y}[1 - P_x^z(t) - P_{y,\vec{x}}^z(t)]\right]^{N_v} \\ &= \left[1 - \frac{2}{V} - \frac{1}{V}\sum_{z,z\neq x,y} P_x^z(t) - \frac{1}{2V}\right. \\ &\quad \left.\times \sum_{z,z\neq x,y}[P_{y,\vec{x}}^z(t) + P_{x,\vec{y}}^z(t)]\right]^{N_v} \\ &= \exp\left(-2\rho_v - \rho_v \sum_{z,z\neq x} P_x^z(t) + \rho_v P_x^y(t)\right. \\ &\quad \left. - \frac{\rho_v}{2}\sum_{z,z\neq x,y}[P_{y,\vec{x}}^z(t) + P_{x,\vec{y}}^z(t)]\right) \end{aligned} \tag{26}$$

where $P_{\vec{x},\vec{y}}^z(t)$ is the probability that a vacancy starts in z at time zero and never reaches either x or y until time t and $P_{x,\vec{y}}^z(t)$ is the probability that a vacancy starts in z at time zero and reaches x at $u \leqslant t$ but never reaches y until time t. In Eqs. (25) and (26) we are left with the calculation of probabilities of the form $P_x^z(t)$, $P_{x,\vec{y}}^z(t) + P_{y,\vec{x}}^z(t)$ for a single random walk. This can be done using Laplace transforms and, concerning $P_x^z(t)$, the computation has been performed a while ago [48]. All the details can be found in Appendix B.

In the continuum limit, $(x-y)/\sqrt{Dt/2} \sim O(1)$; i.e., for independent Brownian motion with diffusion coefficient D, the final expression for $\langle n_x(t)\rangle$ on time scales much larger than one is, in three dimensions,

$$\langle n_x(t)\rangle = \exp[-\rho_v - c_1 D\rho_v t], \tag{27}$$

where c_1 is a constant fixed by the short-length-scale physics—i.e., the underlying lattice structure (see Appendix B). It is clear from this expression which is valid in all di-

mensions larger than 2 that the relaxation time scale is governed by the vacancy density ρ_v and reads $\tau=1/(c_1\rho_v D)$. Physically τ corresponds to the time such that each site has typically been visited once by a defect.

The final expression for G_4 is, for time and length scales much larger than 1, and in the small vacancy density limit $\rho_v \to 0$,

$$G_4(\vec{r},t) = \frac{c_2}{\rho_v}\exp\left(-\frac{2t}{\tau}\right)\left(\frac{t}{\tau}\right)^2 \int_0^1 du \int_0^u dv \frac{e^{-r^2/2Dvt}}{(2\pi Dvt)^{3/2}}, \tag{28}$$

where c_2 is a constant of order unity. Note that the correlation length at fixed t is given by $\xi(t)=\sqrt{Dt}$. For $r \ll \xi(t)$, $G_4(\vec{r},t) \sim 1/r$, whereas for $r \gg \xi(t)$, G_4 decays at leading order as a Gaussian—that is, much faster than exponentially. The $1/r$ behavior is cut off on short-length scales, where Eq. (28) does not hold. For $r=0$ one finds, when $t \gg 1$,

$$G_4(r=0,t) = \langle n_x(t)\rangle - \langle n_x(t)\rangle^2 = \exp(-t/\tau)[1-\exp(-t/\tau)], \tag{29}$$

which behaves as t/τ at small times.

By integrating Eq. (28) over $\vec{r}$ we get the dynamical susceptibility

$$\chi_4(t) = \frac{c_2}{2\rho_v}\left(\frac{t}{\tau}\right)^2 \exp\left(-\frac{2t}{\tau}\right). \tag{30}$$

For short times $t<\tau$, the dynamical susceptibility is proportional to t^2, so that $\mu=2$. This is due to the diffusing nature of the defects. The main contribution to χ_4 is given by the square of the number of sites visited by the same defect, which behaves as $\rho_v(Dt)^2=(1/\rho_v)(t/\tau)^2$, since a random walk in three dimensions typically visits t different sites. For $t>\tau$, on the other hand, the correlation decreases because sites start being visited by different vacancies. The maximum of $\chi_4(t)$ is reached for $t=t^*=\tau$, for which one has $\chi_4(t^*) \sim \rho_v^{-1} \sim Dt^*$. Note that because random walks are fractals of dimension $d_f=2$, the above relation can also be written as $\chi_4(t^*) \sim a^{d-d_f}\xi^{d_f}(t^*)$, where we have added the lattice spacing a to give to χ_4 the dimension of a volume. If for some reason D depends on ρ_v, as happens, for example, for the one-spin-facilitated Fredrickson-Andersen (FA) model where $D \propto \rho_v$, then one finds $t^* \sim \rho_v^{-2}$ and $\chi_4(t^*) \sim t^{*1/2}$.

Taking the Fourier transform of $G_4(r,t)$ given by Eq. (28), we find the four-point structure factor $S_4(k,t)$,

$$S_4(k,t) = \chi_4(t)\mathcal{F}(Dk^2t), \quad \mathcal{F}(u) \equiv \frac{2}{u^2}(u-1+e^{-u}). \tag{31}$$

Note that $S(k=0,t)=\chi_4(t)$, as it should. Furthermore, for large and small k, $S_4(k,t)$ behaves, respectively, as $S_4 \sim k^{-2}$ and $S_4 \sim \chi_4 + O(k^2)$, just as the Ornstein-Zernike form, though the detailed k dependence is different.

One can also study this problem in dimension $d=1$ or $d=2$. Qualitatively, the same conclusions hold (diffusive correlation length $\sqrt{Dt}$, correlation time t^* set by the density of vacancies, etc.), although the quantitative results differ because a random walk in $d \leqslant 2$ visits a number of sites that grows sublinearly with time; see Appendix B 1 and B 3. One finds in particular that $\chi_4(t^*) \sim (Dt^*)^{d/2} \sim \xi^d(t^*)$, with logarithmic corrections for $d=2$. The above arguments can be generalized if for some reason the vacancies have an anomalous diffusion motion, in the sense that their typical excursion between time $t=0$ and time t scales as $t^{1/z}$, where z is the dynamical exponent. When $z=2$, the usual diffusion is observed, but many models like diffusion in random media or kinetically constrained models may lead to subdiffusion, where $z>2$ [21,49]. In this case, one expects the small-time behavior of $\chi_4(t)$ to be given by $\chi_4(t) \sim t^{2d/z}$ for $d<z$ and t^2 for $d>z$ with logarithmic corrections for $d=z$. Similarly, the behavior of $\chi_4(t^*)$ is a power law $\chi_4(t^*) \sim t^{*\lambda}$, with $\lambda=d/z$ for $d<z$ and $\lambda=1$ for $d>z$.

In the above model, mobility defects were assumed to be conserved in time. However, it is certainly more realistic to think that these defects can be spontaneously created and disappear with time. Suppose that defects are created with a rate Γ per unit time and unit volume and disappear with a rate γ per unit time. The equilibrium density of defects is then $\rho_v=\Gamma/\gamma$. The above results on χ_4 can easily be generalized. At small times, the number of pairs of visited sites will now behave as $\rho_v(Dt)^2 - \frac{2}{3}\Gamma(Dt)^3/D$. Because of the death of vacancies, there is an extra decay of the dynamical susceptibility. The dominant rate of decay depends on the adimensional number $\gamma\tau$.

A very similar model for glassy dynamics was suggested in [50], where free volume is described as a diffusing coarse-grained density field $\rho(\vec{r},t)$ with a random Langevin noise term. Mobility of particles is allowed whenever the density ρ exceeds a certain threshold ρ_0. The mobile regions are then delimited by the contour lines of a random field, which already gives rise to a quite complex problem of statistical geometry [51]. The particle density correlation in this model is a simple exponential with relaxation time $\tau \sim \exp(\rho_0/\bar{\rho})$, where $\bar{\rho}$ is the average free-volume density. One can also compute $\chi_4(t)$ in this model to find, in $d=3$,

$$\chi_4(t) \sim t\left\{\exp\left(-\frac{t}{\tau}\right)\left[1-\exp\left(-\frac{t}{\tau}\right)\right]\right\}, \tag{32}$$

which behaves very much like the pointlike vacancy model studied above, with in particular, $\chi_4(t) \sim t^2$ for $t \ll \tau$.

Let us finally note that from the point of view of interacting particles on a lattice we have studied the persistence dynamical susceptibility, instead of the density-density correlations discussed in the Introduction. This is because for the lattice gas problem at hand, the former does not show any interesting properties: except when a defect passes by, the local state is always the same—i.e., occupied. For completeness, we give the corresponding results in Appendix B 4. In a real system, however, the local configuration is going to be affected by the passage of a mobility defect, and one can expect that the density-density correlations will in fact behave more like the persistence dynamical susceptibility computed before. The correspondence between persistence and self-intermediate scattering function is studied explicitly in kinetically constrained models in Ref. [52].

B. Kinetically constrained models: Numerical results

Kinetically constrained models (KCM's) postulate that glassy dynamics can be modeled by forgetting about static interactions between particles, putting all the emphasis on dynamical aspects. Among those models are, for example, the FA model or the Kob-Andersen (KA) model on hypercubic lattices [15,25]. The dynamics of these models can be understood in terms of diffusion of defects [17,21,25] and the models can be classified into cooperative and noncooperative models, depending on the properties of such defects. For cooperative models the size ξ_0, the density, and the time scale for motion of the defects depends on the particle density (for conservative models) or temperature (for nonconservative models) and change very rapidly with increasing density or decreasing temperature [25]. KA and FA models with more than one neighboring vacancy needed in order to allow the motion of other vacancies belong to this class. On the other hand, for the one-spin isotropically facilitated FA model, a single facilitating spin is a mobile defect at all values of temperature and the model is noncooperative. A recent analysis [21] suggests that for these models defects can be considered as noninteracting in $d>4$, while for $d<4$ the role of fluctuations becomes important. Therefore we expect that the previous results for the independent diffusing defects model should apply exactly for FA one-spin facilitated in $d>4$. Furthermore, since the corrections to the Gaussian exponents are not very large [21] in three dimensions, we still expect a semiquantitative agreement. In particular the initial increase of the dynamic susceptibility as $\chi_4(t) \sim N(t)^2$, where $N(t)$ is the total number of distinct visited sites, is expected to be quite a robust result. Also, we expect a diffusive growth of the dynamical length scale $\xi(t)$ governing the scaling of G_4, at least in the limit $\xi(t) \gg \xi_0$. At smaller times, one expects a crossover between a CRR regime when $Dt \ll \xi_0^2$ (where the dynamics inside the defects becomes relevant in cooperative models to a mobility defect regime for longer times). Hence, in principle, looking at the detailed properties of $G_4(r,t)$ one should be able to extract the defect properties—density, size, time scale—and decide which theoretical scenario is most consistent with numerical results.

In the following, we discuss numerical results for the one-spin-facilitated FA model both in $d=1$ and $d=3$ and for the $d=1$ East model where facilitation is anisotropic [15]. The two models can be described, respectively, in terms of diffusive and subdiffusive noncooperative defects and indeed the numerical results are in quantitative agreement with the predictions of the previous section, as will be explained in detail. We do not address the case of cooperative KCM models, for which a more complicated behavior is expected. Indeed a first slowing down of dynamics should occur near a dynamical crossover displaying the properties of an MCT-like avoided transition [25]. In this regime the model cannot be approximated as a system of independent freely diffusing defects and deriving a quantitative prediction for the behavior of four-point correlation and susceptibility would deserve further work. Such avoided transition should then be followed at lower temperature or higher density by an asymptotic behavior described in terms of cooperative diffusing defects.

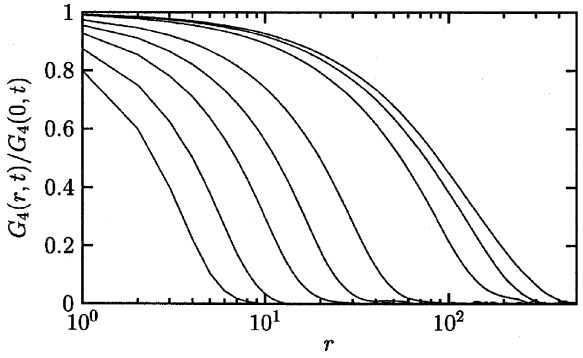

FIG. 2. Four-point spatial correlator (33) in the $d=1$ FA model at fixed temperature $T=0.2$ and various times $t=10^3, 3\times10^3, 10^4, 3\times10^4, 10^5, 10^6, 3\times10^6, 6\times10^6$ (from left to right). The correlator is normalized by its $r=0$ value. At this temperature, the relaxation time is $\tau \sim 10^6$, so that time scales cover both regimes where t/τ is smaller and larger than 1.

1. One dimension

Let us start with the simplest model, the $d=1$ FA model. For a given temperature, we consider the time evolution of the following quantities. The analog of the spatial four-point correlator for this model is

$$G_4(r,t) = \frac{1}{N}\sum_{i=1}^{N} [\langle n_i(t) n_{i+r}(t)\rangle - n^2(t)], \qquad (33)$$

where $n(t)=N^{-1}\Sigma_{i=1}^{N}\langle n_i(t)\rangle$ is the mean persistence, $n_i(t)$ being the persistence at site i. We also measure the corresponding four-point structure factor

$$S_4(k,t) = \frac{1}{N}\sum_{\ell,m=1}^{N} [\langle n_\ell(t) n_m(t)\rangle - n^2(t)] e^{ik\cdot(\ell-m)}, \qquad (34)$$

and as usual we get the four-point susceptibility as the $k \to 0$ limit of the structure factor, $\chi_4(t)=S_4(k=0,t)$. We generally find that the results are in good agreement with the free-defect model described above, at least at sufficiently low temperatures.

In Fig. 2, we show the evolution of the spatial correlator (33) at a given low temperature $T=0.2$ and various times. At this temperature, the relaxation time is about $\tau \sim 10^6$, so that the time scales presented in Fig. 2 cover a range of times both smaller and larger than τ. The dynamic susceptibility $\chi_4(t)$ has the usual shape with a maximum at a time close to τ, indicating that dynamics is maximally heterogeneous there. This nonmonotonic behavior of χ_4 in fact does not show up in the spatial correlators of Fig. 2, which display instead a smooth monotonic evolution with time. The spatial decay of $G_4(r,t)$ becomes slower when t increases, indicating the presence of a monotonically growing dynamic length scale $\xi(t)$.

One can estimate the time dependence of $\xi(t)$ by collapsing the data of Fig. 2 using a form like

$$G_4(r,t) \sim G_4(0,t)\mathcal{G}\left(\frac{r}{\xi}\right). \quad (35)$$

Doing so, we find that $\xi \sim t^{0.45}$ is a reasonable representation of the data at $T=0.2$. Correspondingly, we find that the increase of $\chi_4(t)$ for $t<\tau$ is well described by a power law $\chi_4 \sim t^{0.85}$, so that the expected scaling $\chi_4 \sim \xi^2$ is reasonably verified given the unavoidable freedom in estimating the range of time scales where power laws apply. The values of these exponents are not far from the ones expected from freely diffusing defects in one dimension, although slightly smaller. Indeed, we recall that the results in Appendix B 1 predict $\xi=\sqrt{Dt}$, $\chi_4(t) \propto \rho\xi(t)^2$, and $\chi_4(t^*)=1/\rho$, where ρ is the density of defects, D their diffusion coefficient, and t^* the time at which $\chi_4(t)$ reaches its maximum value. This last prediction is also in good agreement with the numerical results (see, e.g., [20]).

Repeating the simulation at lower temperature $T=0.15$, we obtain $\chi_4 \sim t^{0.93}$, showing that deviations from theoretically expected values are partly due to preasymptotic effects that presumably disappear at very low temperatures.

It is important to remark that the scaling form (35) is only approximately supported by the data. The scaling in fact deteriorates when times become larger than τ. This can be seen in Fig. 2 where data for large times become more and more stretched, indicating an increasing polydispersity of the dynamical clusters. Note that a change in the shape of the spatial correlator makes a quantitative determination of ξ problematic. Usually, one wants to collapse various curves using a form like Eq. (35) to numerically extract ξ. Strictly speaking, this is not possible here if one works at fixed T and varying t over a large time window. This difficulty provides a second possible explanation for the small discrepancy between the measured values of exponents and the theoretical expectations.

The observation of a monotonically growing length begs the question: how can the correlation length increase monotonically with time while the volume integral of the spatial correlator χ_4 is nonmonotonic, as reported in the previous section? This is due to the fact that we have presented in Fig. 2 results for the normalized correlator $G_4(r,t)/G_4(r=0,t)$. By definition, $G_4(0,t)=n(1-n)$; hence, the normalization itself exhibits a nonmonotonic behavior. If one considers the normalized susceptibility $\tilde{\chi}_4=[G_4(0,t)N]^{-1}\Sigma_{\ell,m}[\langle n_\ell n_m\rangle - n^2(t)]$, one indeed finds that $\tilde{\chi}_4$ is monotonically growing as well.

In numerical works, the quantities that have been studied are in fact, most of the time, normalized, and the corresponding $\tilde{\chi}_4(t)$ observed for realistic systems shows a peak, at variance with what is observed in the $d=1$ FA model. As we shall show below, this is due to the one-dimensional nature of the model, and this difference is not observed in three dimensions. This difference in the behavior of the normalized dynamical susceptibility between one and three dimensions is indeed in full agreement with the independent defect diffusion computation; see the previous section and Appendix B.

Results are qualitatively similar in the one-dimensional East model. The dynamic susceptibility $\chi_4(t)$ develops a peak that grows and whose position is slaved to the increasing relaxation time when temperature decreases. At fixed temperature, a monotonically growing length scale is observed, while the scaling relation $\chi_4 \sim \xi^2$ still holds within our numerical precision. The novelty of this model lies in the fact that exponents are now temperature dependent, as all other dynamic exponents in this model. For instance, we find that $\xi(t) \sim t^{0.28}$ at $T=0.4$, $\xi(t) \sim t^{0.15}$ at $T=0.2$. These results are in agreement with the above predictions of the independent defect model if the defect motion is subdiffusive, with a dynamic exponent $z=T_0/T$, as expected from [17]. Due to the quasi-one-dimensional nature of the relaxation process in the three-dimensional generalization of the East model [18], these results most probably carry over to larger dimensions where they would differ by numerical factors only.

2. Three dimensions

In $d=3$, the situation is more subtle. Results for the normalized susceptibility of the one-spin-facilitated FA model were presented in Ref. [22], where it was found to have the standard nonmonotonic shape already described several times above. We find that the non-normalized $\chi_4(t)$ has the same qualitative behavior. Therefore, contrary to the $d=1$ case normalization is not a crucial issue in three dimensions.

In the following we check the predictions for independent diffusing defects in three dimensions for the susceptibility and correlation length obtained above—i.e., $\xi(t)=\sqrt{Dt}$, $\chi_4(t) \propto \rho\xi(t)^4$, and $\chi_4(t^*)=1/\rho$, where ρ is the density of defects, D their diffusion coefficient, and t^* the time at which $\chi_4(t)$ reaches its maximum value. We find a semiquantitative agreement with above prediction, with small deviations in the exponents that should be due to the interaction among defects. In particular the scaling of the peak with the density of defects was already analyzed in [22], where the result $\chi_4(t^*) \propto 1/\rho^{1-\epsilon}$ was obtained, with $\epsilon \simeq 0.03$. As for the correlation length, we find $\xi(t) \propto t^{0.42}$, which shows again a small deviation from the diffusive prediction. Regarding the increase at $t \ll \tau$ of the susceptibility we find a power law as predicted. As in $d=1$, the exponent changes slightly when decreasing temperature because the scaling regime where the power law applies becomes more and more extended. We find $\chi_4 \sim t^{1.4}$ at $T=0.25$, $\chi_4 \sim t^{1.55}$ at $T=0.17$, and $\chi_4 \sim t^{1.89}$ at $T=0.095$. This seems to indicate that the deviation from the scaling $\chi_4(t) \propto t^2$ calculated for the independent diffusing defect model is partly due to preasymptotic effects that are less and less important at lower temperature. Unfortunately, we were not able to measure ξ at much lower temperatures with sufficient accuracy. We expect that even at very low temperature a small deviation from the exponent of independent defects should survive due to the interaction among defects.

In Fig. 3 we show the four-point correlations in both real and Fourier space, Eqs. (33) and (34). In these curves the temperature is fixed at a low value, $T=0.17$, and time is varied in a wide range that includes the relaxation time $\tau(T=0.17) \sim 5\times 10^4$, where the dynamic susceptibility also peaks. For times $t \ll \tau$, the spatial decay of $G_4(r,t)$ is fast. When t increases, the spatial decay becomes slower, once again indicative of an increasing dynamic correlation length

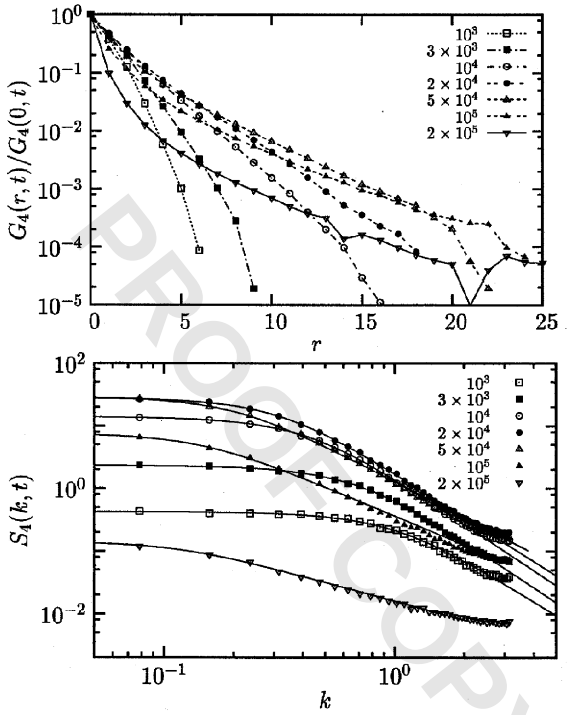

FIG. 3. Four-point correlations in the $d=3$ one-spin-facilitated FA model in both real space (left) and Fourier space (right) at fixed temperature $T=0.17$ and various times indicated in the figures. In Fourier space, points represent numerical data, while solid lines are fits to the form (36) with fitting parameters described in the text.

$\xi(t)$. When t becomes larger than τ, however, spatial correlations seem to become weaker. It is obvious from Fig. 3 that the volume integral of $G_4(r,t)/G_4(0,t)$ decreases when t grows larger than τ. This is very different from the one-dimensional case in Fig. 2, but consistent with all known numerical results.

However, a closer look at Fig. 3 reveals that even though the initial spatial decay of $G_4(r,t)$ is stronger at larger times, the contrary is true at large distances. This indicates that the topology of the dynamic clusters changes when t grows larger than τ, but that $\xi(t)$ may keep increasing in a monotonic manner. Since the spatial correlator is very small at large distances, quantitative measurements of $\xi(t)$ are more easily performed in Fourier space via $S_4(k,t)$.

At short time, a fit of $S_4(k,t)$ using the functional form given by Eq. (31) works reasonably well, but the fit quickly deteriorates at long time. We have therefore used the following generalization of Eq. (31):

$$S_4(k,t) = \chi_4(t)\mathcal{F}_\beta[k^2\xi^2(t)], \quad \mathcal{F}_\beta(u) \equiv \frac{2^{2/\beta}}{u^\beta}(u-1+e^{-u})^{\beta/2}. \tag{36}$$

Freely diffusing defects correspond to $\beta=2$ and $\xi(t)\sim\sqrt{t}$. Using $\beta(t)$ as an additional free parameter, we are able to fit $S_4(k,t)$ at all times; see Fig. 3. We find that β decreases from $\beta\approx2.5$ at small times to $\beta\approx1$ for the longest time scales investigated, which corresponds to $t\approx5\tau$. At such large times, the dynamic susceptibility has already decreased by a factor of ≈300 from its maximum value at $t=\tau$, and correlations become very weak indeed. The values for β found from the fits are consistent with the value $\beta\approx2.15$ reported in Ref. [22] where only fixed time ratio $t/\tau(T)=1$ at different temperatures have been studied. From this fitting procedure, we deduce a monotonically growing dynamic length $\xi(t)$, even beyond $t=\tau(T)$. Fitting its time dependence with a power law, we get $\xi\sim t^{0.42}$ which appears to be slightly subdiffusive, but close to the value found above in the one-dimensional case.

In conclusion we find that on small enough time scales, one indeed has good agreement with the above calculations based on freely diffusing defects; therefore, defect branching and defect coagulation can be neglected. However, for longer time scales, significant deviations appear which correspond to the evolution of the exponent $\beta(t)$ and should be responsible for the small deviations of the predicted exponent for χ_4. Physically, the time evolution of the exponent $\beta(t)$ characterizing the large-k behavior of the dynamic structure factor is reasonable. At very short times, dynamic clusters consist of coils created by random walkers, and an exponent close to $\beta=2$ can be expected. For times $t\sim\tau$, clusters look critical, as described in Refs. [21,22], and the exponent $\beta=2-\eta$, $\eta<0$ is expected. At very large times, clusters are most probably extremely polydisperse because the remaining spatial correlations at large times are due to the largest regions of space that were devoid of defects at time 0 and that take therefore a large time to relax. But at large times, some isolated sites that have not been visited by defects during the relaxation might survive so that the distribution of dynamic clusters at large times is very wide; see Ref. [18] for snapshots. A small value of β can therefore be expected.

VII. NUMERICAL RESULTS ON ATOMISTIC MODEL SYSTEMS

In this section, we study numerical results for the dynamic susceptibility and structure factor of a supercooled liquid simulated by molecular dynamics simulations. The model we study is mainly the well-known binary Lennard-Jones (LJ) mixture as first defined and studied in Ref. [53], but we report also some results for a soft-sphere mixture studied in [38,55,56]. We do not give details about our numerical procedures since these were given several times in the literature [21,30,53].

A. Dynamical susceptibility

In previous works on various realistic liquids, the dynamic susceptibility was reported several times [9,10,12,28]. It is known to exhibit at peak at a time scale enslaved to the quantity chosen to quantify local dynamics. Typically, particle displacements are chosen, and one computes therefore the variance of some dynamical correlation,

$$\chi_4(t) = N[\langle F^2(t)\rangle - \langle F(t)\rangle^2], \tag{37}$$

with

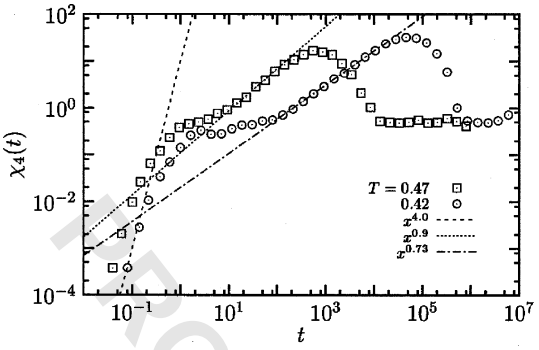

FIG. 4. Time dependence of the dynamic susceptibility in the binary LJ mixture at two different temperatures. The lines are power-law fits with exponents indicated in the label.

$$F(t) = \frac{1}{N}\sum_{i=1}^{N} F_i(t). \tag{38}$$

The dynamic quantity $F_i(t)$ can be chosen as some "persistence" function, in which case $\langle F(t)\rangle$ resembles the overlap function usually measured in spin systems [10,12,28]. Other choices are [21,30]

$$F_i(t) = \cos[\vec{q}\cdot\delta\vec{r}_i(t)], \tag{39}$$

where $\vec{q}$ is a wave vector chosen in the first Brillouin zone and $\delta\vec{r}_i(t)$ is the displacement of particle i in a time interval t. In the limit of small $|\vec{k}|$, it is better to study $F_i(t) = |\delta\vec{r}_i(t)|/\sqrt{\Delta r^2(t)}$, where $\Delta r^2(t)$ is the mean-square displacement of the particles [9,11].

Whereas the general shape of $\chi_4(t)$ is well documented in the literature, its precise time dependence was never discussed. In Fig. 4, we present the time dependence of $\chi_4(t)$ in the binary Lennard-Jones mixture at two different temperatures. The data are presented in a log-log scale, in order to emphasize the existence of several time regimes that are generally hidden in the existing reports. To build these curves, we choose Eq. (39) as the local observable, for a wave vector that corresponds roughly to the typical interparticle distance.

In the ballistic regime at very short times, we find that $\chi_4(t) \sim t^4$, as described in Sec. II. The system then enters the time regime where dynamic structure factors typically exhibit plateaus, as a result of particle caging. As seen in Fig. 4, this is also the case for $\chi_4(t)$. Finally, $\chi_4(t)$ reaches a maximum located close to the relaxation time extracted from the time dependence of $\langle F(t)\rangle$ and then rapidly decays to its long-time limit, equal to $1/2$ in the present case. In Fig. 4, we fitted the time dependence of the increase of $\chi_4(t)$ towards its maximum with power law $\chi_4 \sim t^{\mu}$. The fits are satisfactory, although they only hold on restricted time windows. We find a slight temperature dependence of the exponent μ. For instance, we find $\mu \approx 0.9$ at $T=0.47$ and $\mu \approx 0.73$ at $T=0.42$. As already discussed in the case of kinetically constrained models above, it is not clear how the restricted time window used to determine the exponents might affect their values. However, the data in the Lennard-Jones system behave quantitatively very differently from both theoretical results obtained from freely diffusing defects and numerical results in the one-spin-facilitated $d=3$ FA model, where $\mu=2$. The small temperature evolution in the LJ liquid differs even qualitatively from the one-spin-facilitated $d=3$ FA model where the exponent was found to increase when decreasing temperature. These observations tend to discard a description of this supercooled liquid via a scenario with simple independently diffusing defects, even interacting ones. The above value of μ is in principle compatible with the predictions of elasticity theory, which yields $\mu=1/2$ or $\mu=1$ depending on the damping of phonons. However, the time scale in which the above-mentioned power-law behavior holds in the Lennard-Jones mixture corresponds to the β regime where the displacement of particles is no longer small and the elastic description unjustified. Within MCT, on the other hand, χ_4 should increase in that regime with an exponent $\mu=b$ that is known from previous analysis, $b \approx 0.63$ [54]. The values found above are somewhat larger, but it is hard to know how preasymptotic effects influence the numerical data. Moreover, the value closest to b, $\mu \approx 0.73$, is obtained for $T=0.42$, a temperature already lower than the mode-coupling singularity located at $T_c \approx 0.435$ in this system (a linear interpolation between the values at $T=0.47$ and $T=0.42$ gives $\mu \approx 0.78$ at $T=0.435$). MCT also provides a prediction for the height of the peak, $\chi_4^* \sim t^{*1/\gamma}$, where γ was predicted to be ≈ 2.3, leading to $\lambda = 1/\gamma \approx 0.43$. This prediction is in good agreement with the results of Ref. [21] where $\chi_4(t^*) \sim t^{*0.4}$ was reported. It is important to remark, however, that the MCT exponents are not very well determined. The exponents we reported are the ones computed theoretically in [54]. The exponents obtained from the fits of the numerical data based on MCT are a bit different [53], in particular $b \approx 0.5$ and $1/\gamma \approx 0.37$.

If one insists on using a noncooperative kinetically constrained model to describe the Lennard-Jones liquid, the small value of the short time exponent μ forces one to choose a "fragile" KCM model, such as the East model described above, where the exponent for the dynamic susceptibility is found to be much smaller than the diffusive value $\mu=2$, and indeed to decrease when temperature is decreased. On the other hand, the large dynamic length scales observed in the Lennard-Jones system are not expected for fragile KCM's such as the East model [18]. Our results do not discard the possibility that cooperative KCM's (in a proper density or temperature regime) display a four-point correlation and susceptibility quantitatively similar to the one of the Lennard-Jones liquid. Indeed, as stressed, in e.g., [25], for these models one expects a first regime of slowing down of dynamics due to an avoided mode-coupling transition. The susceptibility and four-point correlation could then well be quantitatively comparable to that of Lennard-Jones liquids. Concerning these comparisons between theoretical scenarios and molecular dynamics simulation results it is important to notice that the the relevance of supposedly "fragile" numerical models for supercooled liquids in shedding light on real fragile glass formers has been questioned [57].

Finally, it is of course a natural question to ask whether the above agreement between MCT predictions and numerical results is only restricted to the Lennard-Jones system. Using the unpublished data of Ref. [56] for a soft-sphere

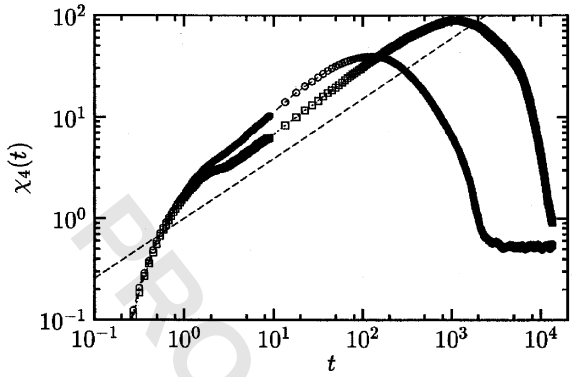

FIG. 5. Dynamic susceptibility $\chi_4(t)$ at $T=0.3$ and 0.26 (from left to right) in a log-log plot as a function of time for the soft-sphere binary mixture of Refs. [38,56]. The data were kindly provided to us by D. Reichman and R. A. Denny. The straight line represents the MCT prediction for the power-law behavior before the peak.

binary mixture where $T_c \simeq 0.22-0.24$ [38,55] we actually found very similar results. Close to T_c a power-law behavior of χ_4 as a function of time can again be observed. For instance, $\chi_4 \sim t^{0.63}$ for $T=0.26$. In Fig. 5 we plot χ_4, defined as in Ref. [28], as a function of time. We also display the power-law behavior predicted by MCT before the peak with the exponent $b \simeq 0.59$ taken from Ref. [55]. There is a similar agreement between the exponent λ measured from the height of the peak and the value of $1/\gamma$ extracted from an MCT analysis of the data. (As in the previous case we used the theoretical MCT exponents computed in [55]. In the case of the soft-sphere system the MCT exponents from numerical fits have probably a large error bars; see the discussion in [54].)

The fact that the predictions of MCT for the four-point susceptibility are in reasonable agreement with numerical simulations in both systems is significant, since the exponents b and $1/\gamma$ are measured on (local) two-point functions and μ and λ on four-point functions. The relation between these exponents test a rather deep structural prediction of MCT that relates time scales to length scales [32]. More numerical work, on other model systems with different values of b, for example, would be needed to establish more firmly whether the coincidence observed in the present paper is or not accidental.

B. Growing length scale?

We focus now more directly on the dynamic length scale. In previous works, the dynamic length scale ξ extracted from four-point correlations was measured either at fixed temperature for various times t where it was found to be nonmonotonic [12,13,27], but monotonic in [11], or at fixed time $t=\tau(T)$, for different temperatures, where it is found to be increasing when the temperature decreases [9,12,21]. In practice, to extract $\xi(t,T)$ from the four-point correlation function either in real space or in Fourier space, one needs to postulate a specific functional form of G_4. In this respect, the results of the previous section on simple lattice KCM's with

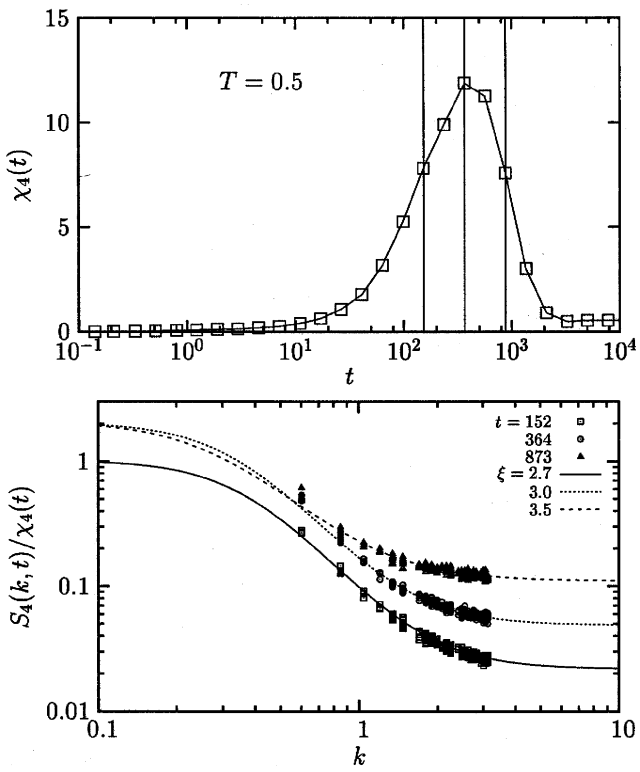

FIG. 6. Left: dynamic susceptibility at $T=0.5$ and $q=4.21$. The vertical lines indicate the times at which $S_4(k,t)$ is evaluated in the bottom figure. Right: the corresponding three $S_4(k,t)$ (the last two have been multiplied by 2 for clarity). Lines are fits to the form (40), the $k \to 0$ limit being fixed by the value of $\chi_4(t)$, with a monotonically growing length scale $\xi(t)$.

no underlying liquid structure prove instructive. It is clear that with data similar to Fig. 3, but obtained with much smaller system sizes, with much less statistics, and polluted by the underlying structure of the liquid, the precise extraction of dynamical length scales from molecular dynamics simulations is not an easy task. More fundamentally, extracting ξ from fitting either $G_4(r,t)$ or $S_4(k,t)$ to a time-independent scaling form necessarily biases the data as discussed above. This also shows that it is a much easier and safer procedure to work, say, at $t=\tau(T)$ and different temperature to observe the growth of a cooperative length $\xi(\tau,T)$ when decreasing T. On the other hand, it is not *a priori* granted that the growth law of ξ with $t=\tau(T)$ when changing T is identical to that of $\xi(t,T)$ with t at a given temperature T. We will not be able to answer this question with our numerical data.

With the above caveats in mind, we present in Fig. 6 some numerical data in the binary Lennard-Jones mixture at a fixed temperature $T=0.5$ and three different times which fall before, at, and after the peak in $\chi_4(t)$. The difficulty of getting clear-cut quantitative determinations for ξ is obvious from Fig. 6. One would need much larger system sizes to properly measure $S_4(k,t)$ at small wave vectors, large times, and low temperatures. The system simulated here contains 1372 particles. One could possibly increase the number of particles by a factor of 10, but the increase in linear size would be very modest, a factor of $10^{1/3} \approx 2.15$. Nonetheless,

we have fitted the data in Fig. 6 with a simple empirical form

$$S_4(k,t) = \frac{\chi_4(t) - C}{1 + (k\xi)^\beta} + C \tag{40}$$

for $0 \leqslant k < k_0$, $k_0 \approx 7.21$ being the position of the first peak in the static structure factor. As for the $d=3$ FA model, the exponent $\beta(t)$ and the dynamic length $\xi(t)$ are fitting parameters. There is an additional free parameter, the additive constant C in Eq. (40), which accounts for the fact that the structure of the liquid starts to be visible and creates some signal in $S_4(k,t)$ when $k \to k_0$. The results of the fitting procedure are presented in Fig. 6 with lines going through the data. Note that the fits in Fig. 6 are constrained at low k by the value of the dynamic susceptibility $\chi_4(t)$. The most important result from Fig. 6 is that if the functional form of $S_4(k,t)$ is given some freedom, here via the time-dependent exponent $\beta(t)$, the extracted dynamic length scale $\xi(t)$ indeed continues to grow monotonically after the peak of the dynamic susceptibility, contrary to reported previously [12,13,27], but in agreement with [11]. We emphasize once more that this result physically makes sense. At times much larger than t^*, only very rare but very large dynamical domains contribute to the dynamic structure factor, so that spatial correlations are weak, but extremely long ranged. The existence of an ever growing length scale is supported by any model with an hydrodynamical limit (such as the phonon or defect models studied here) and is in a sense trivial. The really interesting piece of information is the value of this length scale for $t = \tau_\alpha$—i.e., when the relevant relaxation processes take place.

We conclude that our numerical data are not inconsistent with a monotonically growing length scale even for $t > \tau$, although addressing more quantitative issues such as functional form at the growth law and its temperature dependence would require quite an important, but certainly worthwhile, numerical effort.

VIII. CONCLUSION AND FINAL COMMENTS

Let us summarize the results and various points made in this rather dense paper. First, we have computed numerically and analytically, exactly or approximatively, the four-point correlation function designed to characterize nontrivial cooperative dynamics in glassy systems within several theoretical models: mode-coupling theory, collectively rearranging regions, diffusing defects, kinetically constrained models, and elastic and plastic deformations. The conclusion is that the behavior of $\chi_4(t)$ is rather rich, with different regimes summarized in the Introduction and in Fig. 1. We have computed the early time exponent μ and the peak exponent λ for quite a few different models of glass-forming liquids and shown that the values of these exponents resulting from these models are quite different, suggesting that the detailed study of $\chi_4(t,T)$ should allow one to eliminate or confirm some of the theoretical models for glass formation.

In this spirit, we first simulated some noncooperative KCM's as the one-spin-facilitated FA model in $d=1$ and $d=3$ and the East model. The assumption of pointlike defects that diffuse, possibly with an anomalous diffusion exponent, gives a good account of the shape of the four-point correlation function and of the four-point susceptibility which are in quantitative agreement with the above results for the independent defect model. For strong glasses such as SiO_2, our results might lead to quantitative predictions if the relaxation is indeed due to defect diffusion. It would be very interesting to reconsider numerical simulations of the dynamics of SiO_2 under the light of the present paper to check in more detail that the defect picture is indeed correct in this case (note that our results should enable one to extract, in principle, the properties, density, and relaxation times of defects from the four-point correlation function). For the $d=3$ one-spin-facilitated FA model, we see clear indications of the interactions between defects as time increases. This leads to small deviations of the numerically obtained exponents with respect to those predicted by our analysis of the independent defect model, which does not account for interactions between defects. As far as the identification of a growing length scale $\xi(t)$ from numerical data, we have seen that even within this simplified lattice model, this can be a rather difficult task. Our results point toward a dynamical correlation length that grows forever and a behavior of $S_4(k,t)$ different from the Ornstein-Zernike form but with similar asymptotic behavior. We leave the study of cooperative KCM's, for which a more complicated behavior should occur, for future work. In particular, the detailed form of $S_4(k,t)$ should contain information about the inner structure of the corresponding defects.

We have also analyzed the four-point susceptibility of both a Lennard-Jones system and a soft-sphere system, and shown that the initial exponent μ of the four-point susceptibility is decreasing with the temperature and rather small, $\mu < 1$. We have found, perhaps unexpectedly, a reasonable agreement for μ and λ with the predictions of MCT but not with other theoretical scenarios, such as simple diffusive or subdiffusive defects, strong KCM's, or CRR's (although this might be a question of temperature and time scales, since both CRR and cooperative KCM's are supposed to apply closer to the glass transition temperature). Finally we confirm that the extraction of the growth law of $\xi(t)$ at a given temperature is difficult, and we can only say at this stage that the data are not incompatible with the idea that $\xi(t)$ grows monotonically, even beyond $t = \tau_\alpha$, in the Lennard-Jones system.

As for further work and perspectives, we think that the following points would be worth investigating. First, it would be very interesting to develop a detailed theory of the crossover between the elastic regime described in Sec. III and the mode-coupling β relaxation regime. Is it possible, in particular, to describe approximately the "melting" of the glass as one approaches the mode-coupling transition temperature from below? Second, we only considered systems in equilibrium. One in fact expects that the four-point susceptibility also contains very useful information in the aging regime (see [31,59]). Detailed predictions in this regime may enable one to probe the mechanisms for slow dynamics and the issue of the cooperative length at low temperature in the aging regime [59]. In particular, the elastic contribution

should not age whereas the CRR contribution (characterized by the same exponent μ) should exhibit some aging, possibly allowing one to separate the two effects. Third, since it is clear from the present paper that simpler KCM's (Fredrickson-Andersen one-spin-facilitated, East model and its generalization) seem to fail at describing quantitatively $\chi_4(t)$ obtained by molecular dynamics simulations of (at least two) fragile systems, it would be important to understand if it is possible to find a generalization of these KCM's that can be in agreement with numerics. For the same reason a quantitative study of four-point functions in cooperative KCM's where defects have a complex inner structure would be interesting. Fourth, it would be important to define more complicated correlation functions—for example, a fully general four point function or higher-order correlation functions—in order to test in a more stringent way the idea of cooperativity in glassy systems and distinguish systems where the growth of $\chi_4(t)$ is trivial, such as elastic solids, from those in which a truly nontrivial cooperativity governs the dynamics. Finally, it seems clear that this issue of cooperativity and its associated length scale can only be convincingly settled if long-time scales and low-temperature regimes can be probed quantitatively in experimental systems. We hope that the present paper will motivate ways to directly access four-point functions experimentally in glassy systems (see [31]); natural candidates for this are colloids [3] and granular materials [58,60], although there might be ways to investigate this question in molecular glasses and spin glasses as well [61].

ACKNOWLEDGMENTS

Figure 5 was obtained from the unpublished data of D. R. Reichman and R. A. Denny. We are very grateful to them for providing these results. We thank E. Bertin, D. Chandler, O. Dauchot, J. P. Garrahan, A. Pan, and D. R. Reichman for discussions. G.B. is partially supported by the European Community's Human Potential Programme Contract No. HPRN-CT-2002-00307 (DYGLAGEMEM). C.T. is supported by the European Community's Human Potential Programme Contract No. HPRN-CT-2002-00319 (STIPCO).

APPENDIX A: DYNAMICS OF ELASTIC NETWORKS

1. Four-point correlation function: Overdamped case

We will define $G_4(\vec{r},t)$ for the elastic model defined in the text as

$$G_4(\vec{r},t) = \langle \cos q[\phi(\vec{r},t) - \phi(\vec{r},0)] \times \cos q[\phi(\vec{r}=0,t) - \phi(\vec{r}=0,0)]\rangle - C^2(q,t), \tag{A1}$$

which is equivalent to

$$G_4(\vec{r},t) = \frac{1}{2}\langle \cos q[\phi(\vec{r},t) - \phi(\vec{r},0) + \phi(\vec{r}=0,t) - \phi(\vec{r}=0,0)] + \frac{1}{2}\langle \cos q[\phi(\vec{r},t) - \phi(\vec{r},0) - \phi(\vec{r}=0,t) + \phi(\vec{r}=0,0)] - C^2(q,t).$$

Using the fact that the field ϕ is Gaussian, we finally find

$$G_4(\vec{r},t) = C^2(q,t)\{\cosh[2q^2R(\vec{r},t)] - 1\}, \tag{A2}$$

where

$$R(\vec{r},t) = \langle [\phi(\vec{r},t) - \phi(\vec{r},0)][\phi(\vec{r}=0,t) - \phi(\vec{r}=0,0)]\rangle = \frac{T}{\kappa}\int \frac{d^dk}{(2\pi)^d k^2} e^{-i\vec{k}\cdot\vec{r}}(1 - e^{-\kappa k^2 t}). \tag{A3}$$

Hence,

$$R(\vec{r},t) = \frac{T}{\kappa}(\kappa t)^{1-d/2} F\left(\frac{r}{\sqrt{\kappa t}}\right), \tag{A4}$$

with

$$F(z) = z^{2-d}[I(\infty) - I(z)], \quad I(z) = \int \frac{d^dw}{(2\pi)^d w^2} e^{-iw_1 - w^2/z^2}. \tag{A5}$$

We thus see immediately that $G_4(\vec{r},t)$ will be governed by a "diffusive" correlation length $\xi(t) \sim \sqrt{\kappa t}$, as expected from the structure of the Langevin equation that describes relaxational dynamics. Note that for underdamped dynamics, sound waves would change this scaling.

It is useful to consider the following quantity:

$$J(z) = \frac{\partial I(z)}{\partial\left(\frac{1}{z^2}\right)} = \int \frac{d^dw}{(2\pi)^d} e^{-iw_1 - w^2/z^2}. \tag{A6}$$

In $d=3$, after integrating over dw_1, one has

$$J(z) = \frac{1}{8\pi^{3/2}} z^3 e^{-z^2/4} \tag{A7}$$

and

$$I(z) = \frac{1}{4\pi^{3/2}} \int_z^\infty e^{-u^2/4} du. \tag{A8}$$

Therefore, for $z \ll 1$, one finds $F(z) \simeq (4\pi z)^{-1}$ and $R(\vec{r},t) \simeq T/(4\pi\kappa r)$, whereas for $z \gg 1$,

$$F(z) \simeq (2\pi^{3/2})^{-1} \exp(-z^2/4)/z^2.$$

Thus, for $r \ll \xi(t)$ and $\kappa\Lambda^2 t \gg 1$, the four-point correlation function behaves as

$$G_4(\vec{r},t) = f_q^2\left[\cosh\left(\frac{Tq^2}{2\pi\kappa r}\right) - 1\right]. \tag{A9}$$

2. Four-point correlation function: Underdamped case

We have now

$$m\frac{\partial^2\phi(\vec{r},t)}{\partial^2 t} = \kappa\Delta\phi(\vec{r},t), \tag{A10}$$

which has for solutions in Fourier space

$$\phi_k(t) = \exp(ikVt)\phi_k(0), \tag{A11}$$

with $V=(\kappa/m)^{1/2}$. We now have

$$\langle[\phi(\vec{r},t)-\phi(\vec{r},0)]^2\rangle = \frac{2T}{\kappa}\int\frac{|\exp(ikVt)-1|^2}{k^2}\frac{d^dk}{(2\pi)^d} = \frac{4T}{\kappa}\int(1-\cos[Vkt])dk. \tag{A12}$$

In $d=3$, we find obviously the same result for f_q and G_4 as above, but $R(\vec{r},t)$ is now equal to

$$R(\vec{r},t) = \frac{T}{\kappa}\int\frac{d^dk}{(2\pi)^dk^2}e^{-i\vec{k}\cdot\vec{r}}[1-\cos(kVt)], \tag{A13}$$

which we write

$$R(\vec{r},t) = \frac{T}{\kappa}[I(\vec{r},0)-I(\vec{r},t)], \tag{A14}$$

where

$$I(\vec{r},t) = \int\frac{d^dk}{(2\pi)^dk^2}e^{-i\vec{k}\cdot\vec{r}}\cos(kVt). \tag{A15}$$

By introducing $z=Vt/r$ and changing the variable $q\equiv rk$ and also $u=\cos\theta$ and integrating over u, one finds

$$I(\vec{r},t) = \frac{2\pi}{r}\int dqq^{-1}\{\sin[q(1+z)]+\sin[q(1-z)]\}. \tag{A16}$$

Consider the first term

$$I(\vec{r},t) = \frac{2\pi}{r}\int dqq^{-1}\sin[q(1+z)]. \tag{A17}$$

Changing variable $v=q(1+z)$ directly shows that this integral do not depend on z, as long as $(1+z)$ is positive. This is true for the other integral, which does not depend on z as long as $1-z$ is positive. If $1-z$ is negative, then the integral changes sign. Therefore we have that $I(\vec{r},t)=I(\vec{r},0)$ if $z<1$ and $I(\vec{r},t)=0$ if $z>1$. Therefore $R(\vec{r},t)=0$ if $z<1$ and $R(\vec{r},t)=T/4\pi\kappa r$ when $z>1$. The result is very intuitive: when $z<1$ the information does not have time to travel the distance r and there are no correlation. For $z>1$ the two regions are "connected" and one finds the free-field correlations. Brownian and Newtownian dynamics furnish the same correlation for a given r when the time diverges, as we expect. Finally, it is straightforward to obtain the result quoted in the text for $\chi_4(t)$.

3. Low-dimensional case

We give here, without much detail, the results for elastic networks in $d=1$ and $d=2$. In $d=1$, as is well known, each particle wanders arbitrary far from its initial position but in an anomalous, subdiffusing way, as $t^{1/4}$. Correspondingly, the dynamical structure factor decays as a stretched exponential:

$$\ln C(q,t) \sim \frac{T}{\kappa}q^2t^{1/2}. \tag{A18}$$

Note that the $t^{1/4}$ comes from a collective displacement of the cages and is similar to the anomalous diffusion observed for hard spheres in one dimension, since the latter problem can be mapped onto the Edwards-Wilkinson problem in one dimension [60,62]. We expect that the results obtained here for G_4 should also hold for this case as well. In fact, this model was recently discussed in the context of a simple $d=1$ granular compaction model, see [60].

In $d=2$, the displacement grows logarithmically with time, leading to a power-law decay of the dynamical structure factor with a q-dependent exponent:

$$C(q,t) \sim t^{-y}, \quad y = \frac{q^2T}{8\pi\kappa}. \tag{A19}$$

Turning now to $\chi_4(t)$, we find that after a short transient, $\chi_4(t)$ grows as $t^{1/2}$ in $d=1$ and behaves as t^{1-2y} in $d=2$.

APPENDIX B: CALCULATIONS FOR THE DEFECT MODEL

In Sec. VI we have reduced the computation of $G_4(r,t)$ and $\chi_4(t)$ to probability distributions of a single random walk. In the following we shall show how these quantities can be computed in any spatial dimension.

Let us call $F_x^z(u)$ be the probability that a random walk starting in z reaches x for the first time at time u. $P_x^z(t)$, the probability that a vacancy starts in z at time zero and reaches for the first time x at a time less than t, reads

$$P_x^z(t) = \int_0^t F_x^z(u)du. \tag{B1}$$

Therefore, we need to calculate $F_x^z(u)$. The trick to do that is writing a linear equation relating F_x^z, which we want to compute, to $P^z(x,t)$, the probability that a random walk with self-diffusion coefficient D, starting in z, is in x at time t, which is well known. This linear equation is

$$P^z(x,t) = \delta_{x,z}\delta_{t,0} + \int_0^t F_x^z(u)P^x(x,t-u)du. \tag{B2}$$

By taking the Laplace transform (from now on s is the variable conjugated to t and $\mathcal{L}$ indicates the Laplace transform) we obtain

$$F_x^z(t) = \mathcal{L}^{-1}\left(\frac{\mathcal{L}P^z(x,s)-\delta_{x,z}}{\mathcal{L}P^x(x,s)}\right)(t) \tag{B3}$$

and

$$P_x^z(t) = \int_0^t\mathcal{L}^{-1}\left(\frac{\mathcal{L}P^z(x,s)-\delta_{x,z}}{\mathcal{L}P^x(x,s)}\right)(t')dt' \tag{B4}$$

$$= \mathcal{L}^{-1}\frac{1}{s}\frac{\mathcal{L}P^z(x,s)-\delta_{x,z}}{\mathcal{L}P^x(x,s)}. \tag{B5}$$

A similar strategy can be used to calculate $P^z_{x,\bar{y}}(t)$. Indeed the following equality holds:

$$P^z_{x,\bar{y}}(t) = \int_0^t F^z_{x,\bar{y}}(t')P^x_{\bar{y}}(t-t')dt',$$

$$P^z_{y,\bar{x}}(t) = \int_0^t F^z_{\bar{x},y}(t')P^y_{\bar{x}}(t-t')dt', \tag{B6}$$

where $F^z_{\bar{x},y}(t)$ is the probability that a random walk starting in z at time zero reaches y for the first time at t but never touches x at $s \leqslant t$. Therefore, in order to calculate $P^z_{x,\bar{y}}(t) + P^z_{y,\bar{x}}(t)$ we need to calculate $F^z_{x,\bar{y}}(t) + F^z_{y,\bar{x}}(t)$. It is immediate to check that the following equations hold for any choice of x,z,y:

$$F^z_x(t) = \delta_{x,z}\delta_{t,0} + \int_0^t dsF^z_{\bar{x},y}(s)F^y_x(t-s) + F^z_{x,\bar{y}}(t),$$

$$F^z_y(t) = \delta_{y,z}\delta_{t,0} + \int_0^t dsF^z_{x,\bar{y}}(s)F^x_y(t-s) + F^z_{\bar{x},y}(t), \tag{B7}$$

which implies, again by Laplace transform (z is always different from x and y in the following so we will skip the Kronecker δ's),

$$F^z_{\bar{x},y}(t) + F^z_{x,\bar{y}}(t) = \mathcal{L}^{-1}\frac{\mathcal{L}F^z_x(s) + \mathcal{L}F^z_y(s)}{\mathcal{L}F^x_y(s) + \mathcal{L}F^x_x(s)}. \tag{B8}$$

Using the expression (B3) for $F^x_y(s)$ we get

$$F^z_{\bar{x}}(y,t) + F^z_{\bar{y}}(x,t) = \mathcal{L}^{-1}\frac{\mathcal{L}P^z(x,s) + \mathcal{L}P^z(y,s)}{\mathcal{L}P^x(y,s) + \mathcal{L}P^x(x,s)}. \tag{B9}$$

Furthermore, $P^y_{\bar{x}}(t) = 1 - P^y_x(t)$. Hence we obtain

$$\mathcal{L}P^y_{\bar{x}}(s) = \frac{1}{s} - \mathcal{L}P^y_x(s) = \frac{1}{s}\left(1 - \frac{\mathcal{L}P^y(x,s)}{\mathcal{L}P^x(x,s)}\right). \tag{B10}$$

Finally, we obtain the expression for

$$\mathcal{L}[P^z_{x,\bar{y}}(s) + P^z_{y,\bar{x}}(s)] = \frac{\mathcal{L}P^z(x,s) + \mathcal{L}P^z(y,s)}{\mathcal{L}P^x(y,s) + \mathcal{L}P^x(x,s)}\frac{1}{s}\left(1 - \frac{\mathcal{L}P^y(x,s)}{\mathcal{L}P^x(x,s)}\right). \tag{B11}$$

A useful way to rewrite this expression is obtained by summing and subtracting the Laplace transform of $P^z_x(t) + P^z_y(t)$:

$$P^z_{x,\bar{y}}(t) + P^z_{y,\bar{x}}(t) = P^z_x(t) + P^z_y(t) - 2\mathcal{L}^{-1}\frac{\mathcal{L}P^z(x,s) + \mathcal{L}P^z(y,s)}{\mathcal{L}P^x(y,s) + \mathcal{L}P^x(x,s)}\frac{1}{s}\frac{\mathcal{L}P^y(x,s)}{\mathcal{L}P^x(x,s)}. \tag{B12}$$

Finally putting together all the different terms we have

$$\langle n_x(t)n_y(t)\rangle = \exp[-2\rho_v - 2\rho_v N(t) + 2\rho_v P^y_x(t) + \rho_v G(t,x-y)], \tag{B13}$$

where $N(t) = \Sigma_{z\neq x}P^z_x(t)$ is the average number of distinct sites (minus 1) visited by a random walk during the interval of time t and

$$G(t,x-y) = \mathcal{L}^{-1}\left[\sum_{z\neq x,y}\frac{\mathcal{L}P^z(x,s) + \mathcal{L}P^z(y,s)}{\mathcal{L}P^x(y,s) + \mathcal{L}P^x(x,s)}\frac{1}{s}\frac{\mathcal{L}P^y(x,s)}{\mathcal{L}P^x(x,s)}\right]. \tag{B14}$$

Since

$$\langle n_x(t)\rangle^2 = \exp[-2\rho_v - 2\rho_v N(t)], \tag{B15}$$

the expression of G_4 is

$$G_4(x-y,t) = \exp[-2\rho_v - 2\rho_v N(t)] \times\{\exp[2\rho_v P^y_x(t) + \rho_v G(t,x-y)] - 1\}. \tag{B16}$$

In the following we shall analyze separately the one-dimensional case, the three- or higher-dimensional case, and the two-dimensional case.

1. One dimension

Consider a symmetric random walk on a one-dimensional lattice with lattice spacing a. By Laplace transforming the master equation

$$\frac{dP^z(x,t)}{dt} = \frac{P^z(x+a,t) + P^z(x-a,t) - 2P^z(x,t)}{2}, \tag{B17}$$

one immediately obtains

$$\mathcal{L}P^z(x,s) = \int_{-\pi/a}^{\pi/a}\frac{dk}{2\pi}\frac{e^{ik(x-z)}}{\zeta(k)+s}, \tag{B18}$$

where $\zeta(k) = (1-\cos k)$. In the continuum limit $a\to 0$, $(x-y) \propto a\sqrt{Dt/2} \propto a^2$, the above integral can be solved with the well-known result

$$\mathcal{L}P^z(x,s) = \frac{1}{\sqrt{4Ds}}e^{-\sqrt{s}|x-z|/\sqrt{D}}, \tag{B19}$$

which correspond to the solution of the diffusion equation for a one-dimensional Brownian motion with diffusion coefficient D—i.e.,

$$\frac{dP}{dt} = D\frac{d^2P}{dx^2}. \tag{B20}$$

Let us now compute all the functions needed to get G_4. First,

$$N(t) = \sum_{z\neq x}P^z_x(t) = \sum_{z\neq x}\mathcal{L}^{-1}\left(\frac{1}{s}\frac{\mathcal{L}P^z(x,s)}{\mathcal{L}P^x(x,s)}\right)(t),$$

where we used Eq. (B4). When $t \gg 1$ we get

$$N(t) = 4\frac{\sqrt{Dt}}{\sqrt{\pi}}.$$

Second, using the expression (B14) of G in terms of $\mathcal{L}P^z(x,s)$ we get

$$\mathcal{L}G(s,x-y) = 2\frac{\sqrt{D}}{s^{3/2}}\frac{e^{-\sqrt{s}|x-y|/\sqrt{D}}}{e^{-\sqrt{s}|x-y|/\sqrt{D}}+1}.$$

Changing variable in the inverse Laplace transform we get

$$G(t) = 4\sqrt{Dt}f\left(\frac{|x-y|}{\sqrt{2Dt}}\right),$$

where $f((|x-y|)/\sqrt{2Dt})$ equals

$$f\left(\frac{|x-y|}{\sqrt{2Dt}}\right) = \int_{-i\infty-\gamma}^{+i\infty-\gamma}\frac{e^{-\sqrt{2s}|x-z|/\sqrt{Dt}}}{e^{-\sqrt{2s}|x-z|/\sqrt{Dt}}+1}e^{-s}\frac{ds}{s^{3/2}}.$$

Finally $P_x^y(t)$ can be computed easily but it is always much smaller than the other terms in the exponential, so we are going to neglect it. The resulting expression for G_4 is

$$G_4(x-y,t) = \exp\left(-2\rho_v - \frac{8\rho_v}{\sqrt{\pi}}\sqrt{Dt}\right) \times\left\{\exp\left[\rho_v 2\sqrt{Dt}f\left(\frac{|x-y|}{\sqrt{2Dt}}\right)\right]-1\right\}. \quad \text{(B21)}$$

Note that the typical time scale is $\tau = 1/\rho_v^2 D$, and since we focus on $\rho_v \to 0$, we can rewrite the above expression as

$$G_4(x-y,t) = \exp\left(-\frac{\sqrt{8}}{\sqrt{\pi}}\sqrt{t/\tau}\right) \times\left\{\exp\left[2\sqrt{t/\tau}f\left(\rho_v\frac{|x-y|}{\sqrt{2t/\tau}}\right)\right]-1\right\}. \quad \text{(B22)}$$

Integrating over $x-y$ to get the χ_4 we find

$$\chi_4(t) = \frac{2}{\rho_v}\exp\left(-\frac{8}{\sqrt{\pi}}\sqrt{t/\tau}\right)\sqrt{2t/\tau}\int_0^{+\infty}dx\{\exp[2\sqrt{t/\tau}f(x)]-1\}. \quad \text{(B23)}$$

In particular when $t/\tau \ll 1$ we have

$$\chi_4(t) \propto \frac{1}{\rho_v}(t/\tau). \quad \text{(B24)}$$

The interpretation of this result is that at short times the defects do not intersect and the χ_4 is just the square of the number of average sites visited by a random walk until time t. We will see that this interpretation is indeed correct in any dimension.

Finally, after some algebra it is possible to obtain from Eq. (B24) that $\chi_4(t) \simeq (c/\rho_v)\exp[(-4/\sqrt{\pi})\sqrt{t/\tau}]$ at very large times (c is a numerical constant). Thus, as found in simulations, the normalized χ_4 does not go to zero as it happens in three dimensions.

2. Three dimensions and higher

Consider a symmetric random walk on a cubic lattice. The general expression for $P^z(x,s)$ is

$$P^z(x,s) = \int_{BZ}\frac{d^dk}{(2\pi)^d}\frac{e^{ik(x-z)}}{\zeta(k)+s}, \quad \text{(B25)}$$

where BZ means Brillouin zone and $\zeta(k) = \Sigma_{i=1}^d(1-\cos k_i)$ for a hypercubic lattice (k_i is the component of $\vec{k}$ in the direction i). Also in this case we consider the continuum limit $(x-y)/\sqrt{Dt/2} \propto O(1)$ and look for times t much larger than 1.

Let us again compute all the needed quantities: first, $N(t)$. In this case for $t \gg 1$ we find that

$$N(t) = D\left(\int_{BZ}\frac{d^dk}{\pi\zeta(k)}\right)^{-1}\mathcal{L}^{-1}\frac{1}{s^2}.$$

Hence $N(t) = c_1 tD$ where $c_1 = [\int_{BZ}d^dk/\pi\zeta(k)]^{-1}$.

Again, we neglect the $P_x^y(t)$ term and we focus on G in the continuum limit, for $t \gg a$. We get

$$\mathcal{L}G = \frac{1}{s^2}\frac{\int_{BZ}\frac{d^dk}{(2\pi)^d}\frac{e^{ik(x-z)}}{Dk^2+s}}{\left(\int_{BZ}\frac{d^dk}{(2\pi)^d}\frac{1}{Dk^2}\right)^2}.$$

Changing variable in the inverse laplace transform we get

$$G(t) = D^2\int_{-i\infty-\gamma}^{+i\infty-\gamma}e^{ts}\int_{BZ}\frac{d^dk}{(2\pi)^d}\frac{e^{ik(x-z)}}{Dk^2+s}\frac{\exp(ts)}{C^2s^2}ds.$$

Since we know the inverse laplace transform of the function resulting from the integral over k [it is simply $P^y(x,t)$] and each $1/s$ adds an integral, we finally get

$$G(t) = c_2(Dt)^2\int_0^1 du\int_0^u dv\frac{e^{-(x-y)^2/2Dtv}}{(2\pi Dtv)^{d/2}},$$

where c_2 is a numerical constant of order unity. From this expression, we finally obtain

$$G_4(x-y,t) = \exp(-2\rho_v - 2\rho_v c_1 Dt)\left[\exp\left(\rho_v(c_2Dt)^2 \times\int_0^1 du\int_0^u dv\frac{e^{-(x-y)^2/2Dtv}}{(2\pi Dtv)^{3/2}}\right)-1\right] \quad \text{(B26)}$$

and the results quoted in the main text.

3. Two dimensions

In two dimensions things are a bit tricky because of logarithmic corrections. Briefly, we obtain that

$$G_4(x-y,t) = \exp\left(-2\frac{c_3 t}{\tau\ln t}\right)\frac{1}{\rho_v}c_4^2(t/\tau)^2\frac{1}{(\ln tD)^2}\int_0^1 du \times\int_0^u dv\frac{e^{-(x-y)^2/2Dvt}}{(2\pi Dvt)}, \quad \text{(B27)}$$

with c_3 and c_4 constants of order unity. Hence, integrating over $x-y$, we get

$$\chi_4(t) = \exp\left(-2\frac{c_3 t}{\tau\ln t}\right)\frac{1}{2\rho_v}c_4^2(t/\tau)^2\frac{1}{(\ln tD)^2}. \quad \text{(B28)}$$

In particular when $t/\tau \ll 1$ we have

$$\chi_4(t) \propto \frac{1}{\rho_v}\left(\frac{t}{\tau \ln t}\right)^2. \tag{B29}$$

Again, since the number of sites visited on average by a RW in two dimensions goes like $t/\ln t$, at short times χ_4 is the square of the number of average sites visited until time t.

4. Density-density correlations

We now sketch the calculation for the density four point correlation, defined as

$$G_4^d(x-y,t) \equiv \langle [\eta_x(t)\eta_x(0)-\rho^2][\eta_y(t)\eta_y(0)-\rho^2]\rangle - \langle \eta_x(t)\eta_x(0)\rangle_c^2, \tag{B30}$$

with $\eta_x(t)=0,1$ if the site x is empty or occupied at time t, respectively. We start from

$$\langle \eta_x(t)\eta_x(0)\rangle_c^2 = \left(\left[\frac{1}{V}\sum_{z,z\neq x}[1-P^z(x,t)]\right]^{N_v} - \rho^2\right)^2.$$

Using that $\Sigma_z P^z(x,t)=1$ we get

$$\langle \eta_x(t)\eta_x(0)\rangle_c^2 = \exp(-4\rho_v)\{\exp[\rho_v P^x(x,t)]-1\}^2. \tag{B31}$$

In the limit $\rho_v \to 0$ we have

$$\langle \eta_x(t)\eta_x(0)\rangle_c^2 = [\rho_v P^x(x,t)]^2. \tag{B32}$$

Similarly we find that

$$\langle \eta_x(t)\eta_x(0)\eta_y(t)\eta_y(0)\rangle = \left(\frac{1}{V}\sum_{z,z\neq x,y}[1-P^z(x,t)-P^z(y,t)]\right)^{N_v} = \exp[-4\rho_v + 2\rho_v P^x(x,t) + 2\rho_v P^y(x,t)].$$

Collecting all the pieces together we finally get, at leading order in ρ_v,

$$G_4^d(x-y,t) = 2\rho_v P^y(x,t) \tag{B33}$$

for $x \neq y$. The interpretation of this equation is that the dynamical correlation between x and y is due to the fact that the *same vacancy* was in x at time 0 and t at time t or vice versa. Integrating over $x-y$ one finds that at long times $\chi_4(t) \propto 1/t^{d/2}$, showing no interesting structure.

[1] G. Adam and J. H. Gibbs, J. Chem. Phys. **43**, 139 (1965).
[2] See, e.g., M. D. Ediger, Annu. Rev. Phys. Chem. **51**, 99 (2000).
[3] E. Weeks, J. C. Crocker, A. C. Levitt, A. Schofield, and D. A. Weitz, Science **287**, 627 (2000).
[4] R. Richert, J. Phys.: Condens. Matter **14**, R703 (2002).
[5] E. Vidal-Russell and N. E. Israeloff, Nature (London) **408**, 695 (2000).
[6] L. A. Deschenes and D. A. Vanden Bout, Science **292**, 255 (2001).
[7] M. M. Hurley and P. Harrowell, Phys. Rev. E **52**, 1694 (1995) and references therein.
[8] A. Widmer-Cooper, P. Harrowell, and H. Fynewever, Phys. Rev. Lett. **93**, 135701 (2004).
[9] R. Yamamoto and A. Onuki, Phys. Rev. Lett. **81**, 4915 (1998) and references therein.
[10] G. Parisi, J. Phys. Chem. B **103**, 4128 (1999).
[11] B. Doliwa and A. Heuer, Phys. Rev. E **61**, 6898 (2000).
[12] C. Bennemann, C. Donati, J. Baschnagel, and S. C. Glotzer, Nature (London) **399**, 246 (1999); S. C. Glotzer, V. V. Novikov, and T. B. Schrøder, J. Chem. Phys. **112**, 509 (2000).
[13] N. Lacevic, F. W. Starr, T. B. Schrøder, V. N. Novikov, and S. C. Glotzer, Phys. Rev. E **66**, 030101(R) (2002); N. Lacevic and S. C. Glotzer, J. Phys.: Condens. Matter **15**, S2437 (2003); N. Lacevic, F. W. Starr, T. B. Schrøder, and S. C. Glotzer, J. Chem. Phys. **119**, 7372 (2003).
[14] Y. Hiwatari and T. Muranaka, J. Non-Cryst. Solids **235–237**, 19 (1998).
[15] F. Ritort and P. Sollich, Adv. Phys. **52**, 219 (2003).
[16] J. Jäckle and S. Eisinger, Z. Phys. B: Condens. Matter **84**, 115 (1991).
[17] P. Sollich and M. R. Evans, Phys. Rev. Lett. **83**, 3238 (1999).
[18] L. Berthier and J. P. Garrahan, e-print cond-mat/0410076.
[19] G. H. Fredrickson and H. C. Andersen, Phys. Rev. Lett. **53**, 1244 (1984); J. Chem. Phys. **83**, 958 (1984).
[20] J. P. Garrahan and D. Chandler, Phys. Rev. Lett. **89**, 035704 (2002); Proc. Natl. Acad. Sci. U.S.A. **100**, 9710 (2003).
[21] S. Whitelam, L. Berthier, and J. P. Garrahan, Phys. Rev. Lett. **92**, 185705 (2004).
[22] S. Whitelam, L. Berthier, and J. P. Garrahan, e-print cond-mat/0408694.
[23] L. Berthier and J. P. Garrahan, J. Chem. Phys. **119**, 4367 (2003); Phys. Rev. E **68**, 041201 (2003).
[24] W. Kob and H. C. Andersen, Phys. Rev. E **48**, 4364 (1993).
[25] C. Toninelli, G. Biroli, and D. S. Fisher, Phys. Rev. Lett. **92**, 185504 (2004); e-print cond-mat/0410647.
[26] ■. ■■■■■, J. Chem. Phys. **33**, 1371 (1960).
[27] A. Pan, D. Chandler, and J. P. Garrahan, e-print cond-mat/0410525.
[28] S. Franz and G. Parisi, J. Phys.: Condens. Matter **12**, 6335 (2000); C. Donati, S. Franz, G. Parisi, and S. C. Glotzer, J. Non-Cryst. Solids **307**, 215 (2002).
[29] T. R. Kirkpatrick and D. Thirumalai, Phys. Rev. A **37**, 4439 (1988).
[30] L. Berthier, Phys. Rev. E **69**, 020201(R) (2004).
[31] P. Mayer, H. Bissig, L. Berthier, L. Cipelletti, J. P. Garrahan, P. Sollich, and V. Trappe, Phys. Rev. Lett. **93**, 115701 (2004).
[32] G. Biroli and J.-P. Bouchaud, Europhys. Lett. **67**, 21 (2004).
[33] L. Berthier, Phys. Rev. Lett. **91**, 055701 (2003).
[34] T. Kirkpatrick and P. Wolynes, Phys. Rev. B **36**, 8552 (1987); T. R. Kirkpatrick, D. Thirumalai, and P. G. Wolynes, Phys. Rev. A **40**, 1045 (1989).
[35] J.-P. Bouchaud and G. Biroli, J. Chem. Phys. **121**, 7347 (2004).

[36] G. Tarjus and D. Kivelson, J. Chem. Phys. **103**, 3071 (1995); D. Kivelson and G. Tarjus, Philos. Mag. B **77**, 245 (1998).
[37] W. Götze and L. Sjögren, Rep. Prog. Phys. **55**, 241 (1992); W. Götze, Condens. Matter Phys. **1**, 873 (1998); W. Kob, in *Slow Relaxations and Non-equilibrium Dynamics in Condensed Matter*, edited by J.-L. Barrat, M. Feigelman, J. Kurchan, and J. Dalibard (Springer, Berlin, 2003).
[38] T. S. Grigera, A. Cavagna, I. Giardina, and G. Parisi, Phys. Rev. Lett. **88**, 055502 (2002).
[39] A. Montanari and G. Semerjian, e-print cond-mat/0412023.
[40] M. Vogel, B. Doliwa, A. Heuer, and S. C. Glotzer, J. Chem. Phys. **120**, 4404 (2004).
[41] L. V. Woodcock, C. A. Angell, and P. Cheeseman, J. Chem. Phys. **65**, 1565 (1976).
[42] S. A. Brawer, Phys. Rev. Lett. **46**, 778 (1981); J. Chem. Phys. **75**, 3516 (1981); D. A. Litton and S. H. Garofalini, J. Non-Cryst. Solids **217**, 250 (1997).
[43] See, e.g., C. Masciovecchio, G. Ruocco, F. Sette, M. Krisch, R. Verbeni, U. Bergmann, and M. Soltwisch, Phys. Rev. Lett. **76**, 3356 (1996).
[44] Note, however, that in a numerical simulation with a box of size, say, $L=10a$, the elastic contribution will saturate after a time $L/V\sim 1$ ps. This contribution will therefore be hard to see unless much larger sizes can be simulated.
[45] G. Biroli, L. Berthier, and J.-P. Bouchaud (unpublished).
[46] D. S. Fisher and D. A. Huse, Phys. Rev. B **38**, 373 (1988).
[47] S. Redner, *A Guide to First Passages Processes* (Cambridge University Press, Cambridge, U.K., 2001).
[48] *Fluctuation Phenomena Studies in Statistical Mechanics*, Vol. 7 of edited by E. W. Montroll and J. L. Lebowitz (North-Holland, Amsterdam, 1979).
[49] E. Bertin, J. P. Bouchaud, and F. Lequeux, e-print cond-mat/0501192.
[50] J.-P. Bouchaud, in *Slow Relaxations and Non-equilibrium Dynamics in Condensed Matter*, edited by J.-L. Barrat, M. Feigelman, J. Kurchan, and J. Dalibard (Springer, Berlin, 2003).
[51] M. B. Isichenko, Rev. Mod. Phys. **64**, 961 (1992).
[52] L. Berthier, D. Chandler, and J. P. Garrahan, Europhys. Lett. **69**, 320 (2005).
[53] W. Kob and H. C. Andersen, Phys. Rev. Lett. **73**, 1376 (1994).
[54] M. Nauroth and W. Kob, Phys. Rev. E **55**, 657 (1997).
[55] J.-L. Barrat and A. Latz, J. Phys.: Condens. Matter **2**, 4289 (1990).
[56] D. R. Reichman and R. A. Denny (unpublished).
[57] G. Tarjus, D. Kivelson, and P. Viot, J. Phys.: Condens. Matter **12**, 6497 (2000).
[58] G. Marty and O. Dauchot, Phys. Rev. Lett. **94**, 015701 (2005).
[59] H. Castillo, C. Chamon, L. F. Cugliandolo, J. L. Iguain, and M. P. Kennett, Phys. Rev. B **68**, 134442 (2003); C. Chamon, P. Charbonneau, L. F. Cugliandolo, D. R. Reichman, and M. Sellitto, J. Chem. Phys. **121**, 10120 (2004).
[60] A. Lefèvre, L. Berthier, and R. Stinchcombe, e-print cond-mat/0410741.
[61] J. P. Bouchaud and G. Biroli, e-print cond-mat/0501668.
[62] R. Arratia, Z. Ann. Prob. **11**, 382 (1983).

Abstract

The price fluctuations of the stock market display fascinating properties. The volatility is around one order of magnitude too large than what is predicted by the efficient market theory, and is correlated on very large time scales. Agents overreact to new pieces of information. We show that such properties spontaneously appear when agents use their experience and the past behavior of the market to take decisions. We also study the price formation and the microstructure of financial markets, at the level of the order book which organizes transactions. We explain why the price is diffusive despite the fact that market orders (the shocks that impact price) are long-range correlated in time. We evaluate the spread using symmetry arguments, and show that it is directly proportional to the volatility per trade.

Part II: Price fluctuations, Conventions and Microstructure of Financial Markets

December 6, 2005

Contents

1. Introduction

As statistical physics, economics aims at describing the equilibrium and dynamics of a large number of entities, such as agents or companies, that are interacting with each other. In both cases these interactions can lead to sudden collective phenomena. In economics, crashes occur where agents act collectively and sell their stocks. Sometimes there is no obvious external reason for such event, as in October 1987 where the stock market lost more than 20 percents in one single day. Collective behavior is also apparent during bubbles, or to explain many fads and fashions that affect markets at large, with a much wider scope than financial markets.

One of the main goal of economics is to understand the decision making problem: how do agents take decisions. Neoclassical theory posits that agents are fully rational (Mirowski, 1989). It is much debated today whether the market behavior, for example the presence of crashes or bubbles can be well understood in this framework (Shiller, 2000, 2002; Shleifer, 2000). Furthermore the neoclassical method itself has been criticized (White, 2004): whenever empirical data differs from the equilibrium predictions of the neoclassical theory (the so-called 'anomalies'), extra degrees of freedom are added, sometimes without parsimony, to account for these variations. Among others, the "econophysics" approach is not axiomatical, but empirical, and aims to rationalize the empirical data with parsimonious (even if sometimes unrealistic) models.

The stock market is a very convenient setting to study how agents take decisions, firstly because agents are supposed to be highly concerned and to act reasonably when money is involved, and secondly because recently the quantity of available data has become huge. The most direct observable is the price. Below a certain time scale, which depends on the market considered (for liquid markets it is of the order of the day), the microscopic organization of the market, the 'microstructure', plays an important role. The two regimes (short times and longer times) reflect interesting aspects of the behavior of agents. In what follows we study both, respectively in Chapters 2 and 3. In the present chapter we introduce the two topics, and we start with the long time regime.

1.1 Price fluctuations

One hundred years ago, a few years before Einstein introduced it to explain the Brownian motion, the random walk theory was established by Bachelier to describe price fluctuations. The fact that *properly anticipated prices fluctuate randomly*, which was already quite understood by Bachelier, was latter justified by Samuelson (Samuelson, 1965). The explanation given was as follows: in the efficient market theory (Fama, 1970), where agents are fully rational, the price is fair, or "efficient": it represents perfectly the fundamental reality of the economy. In particular the price contains all the informations available on the market: when it is not the case, smart agents can use these informations to forecast the price. They can buy or sell according to their forecast, which moves the price to its fair value. Hence price changes can only correspond to new informations. By definition, new informations are de-correlated from the past, therefore the price must fluctuate randomly. Empirically, temporal correlations in the sign of the price fluctuations are rather small, at least on average, as we shall see in Chapters 2 and 3. Nevertheless this does not support that prices are efficiency and equal to their true, fundamental value, as simpler mechanism can destroy possible correlations [1]. Furthermore, there are important differences between the behavior of price fluctuations and a classical random walk. All of them should be explained by any eligible theory of financial markets: (i) The distribution of absolute price changes $x = |P_t - P_0|$ for small time (say less than one day) is highly non Gaussian. Its tails are well described by a Pareto law $P(x) \sim x^{-1-\mu}$, where the value $\mu \approx 3$ seems to be universal. The far tail of this distribution corresponds to the crashes mentioned above. (ii) The diffusion constant of the price, the volatility, is itself a random variable that displays very long, power law correlations in time, analogous to fluctuations in turbulent flows. This reflects the intermittent nature of the market that displays bursts of activities and calm periods at many different time scales. This effect is called *volatility clustering* (iii) In a famous study Shiller showed that the volatility is at least a factor of 5 too large compared to what is expected in the efficient market theory (Shiller, 1981) (iv) the market over-reacts to new pieces of information (de Bondt and Thaler, 1985).

In Chapter 2 we aim to relate some of these facts to the problem of decision making of agents, and to the interactions that exist between them.

[1] In particular, the 'smart agents' could simply be arbitrageurs who take decisions only on the past price fluctuations, and not on the fundamental value of the stocks. This is the behavior of most of the hedge funds, who represents a non-negligible fraction of the trades in financial markets. If correlations were there, simple strategies could use them to make profit. These strategies would diminish these correlations.

We shall propose a simple model (Wyart and Bouchaud, 2003) close in spirit to a well known metaphor of Keynes. The market is compared to a beauty contest where the players, instead of voting for the girl they prefer, try to vote for the one who will be elected. This emphasizes that some participants are more interested to guess the opinion of others than to discover the fundamental value of a stock. Our model describes agents who try to foresee how a new piece of information (that could be a financial index, an economic indicator, or the past price itself) will impact the price. To do so, they study the past statistical correlations between this information and the price. Then they use this correlation to devise active trading strategies. Obviously, the impact of these strategies on the price modifies the observed correlations. This positive feedback generates both excess volatility and over-reaction to information. Furthermore, when the proportion of these active agents increases, the system undergoes a phase transition and two stable states arise. Such states are examples of *self fulfilling prophecies* (Woodford, 1990), or conventions (Orléan, 1999). The dynamics of the market is then dominated by rare switches between these conventions, where the correlations change value. Empirical evidences show that such dynamics does indeed take place; a striking example is the correlation between bonds and equities that suddenly changed *sign* in the 90's as we shall document here. This dynamics also generate long term correlations in the *volatility* as observed in financial markets. This illustrates that some of the rules that govern the market can *evolve* in time, in stark contrast with the predictions of the efficient market theory. As these rules are not universal, they can also differ between continents: employment affects the US market more than in Europe, where inflation is the prevailing information (Briere, 2005).

1.2 Microstructure

At short time scales, say of the order of the day, the price fluctuations depend on the microscopic organization of the market, the so-called 'microstructure'. How much the microstructure details affect prices is an important practical question. In particular, the microstructure might influence the overall market volatility. This idea is at the center of the Tobin tax, which posits that increasing the transaction costs will diminish speculation, and will therefore lower the volatility. The microstructure is also of primordial theoretical interest, since it rules how offer and demand interact to form a price, a basic question in economics. In what follows we will be particularly interested in one type of microstructure: "continuous double-auction", which is the standard mechanism of the price formation in most modern electronic financial

markets. We shall detail the organization of electronic markets in Chapter 3. It is a completely self-organized process, where all agents act under the same rules. In particular there are no official market makers (even though some agents may act as market makers and furnish liquidity). A continuous double-auction is organized with an *order-book*, that makes the inventory of all available offer and demand, as we shall detail more in Chapter 3. Offer and demand are expressed by *limit orders*. These are propositions made by agents, the "liquidity providers", to buy or sell stocks at a price they fix. At any given time there is a best (lowest) offer, or ask, to sell with price $a(t)$, and a best (highest) bid to buy at price $b(t)$. Transaction occurs when *market orders* are emitted by impatient agents, the "liquidity takers". The price of the transaction is $a(t)$ if the market order is to buy, $b(t)$ if the market order is to sell. Sketches of order-books are presented in Fig.(3.1), and empirical average of half order-books are shown in Fig.(1.3).

Only recently a huge amount of data became available on this topic. Several empirical works were done to observe if universal laws emerge from this extremely strategic, self-organized world. An important micro-structural feature is the market *liquidity*, which quantifies transaction costs. Liquidity is characterized by the *spread* $S(t) = a(t) - b(t)$, which is the cost associated with the buying, and the following instantaneous selling of one share. The spread can vary by one order of magnitude between different stocks. There is yet no accepted theory on what fixes it. Recently, empirical data of stocks with small "ticks" (the tick is the minimal increment of price, see Chapter 3) show that the spread depends linearly with the volatility *per trade*, which is the root mean square of the price increment between two trades (Zumbach, 2004; Kockelkoren et. al, 2005). New results are shown for France-Telecom in Fig1.1. There is not yet explanation for this fundamental law. This empirical relation was not observed before in the microstructure litterature, where the correlation between spread and volatility *per unit time* is considered in general. This leads to inconclusive results (Pastor & Stambaugh, 2003; Chordia et al., 2001).

Another characteristic of the liquidity is the *price impact function* $\mathcal{R}(V)$, which is the average change of price following the buying of a volume V of shares. It was observed that when the volume and the price impact are properly rescaled by some powers of the stock capitalization, the price impact functions of the different stocks of the New York Stock Exchange collapse on a single curve (Lillo, Farmer and Mantegna 2002), as shown in Fig.(1.2). Somewhat surprisingly, this curve is strongly concave, and it can be reasonably well fitted by a logarithm (Potters and Bouchaud, 2003) or by two power laws with small exponents (Lillo, Farmer and Mantegna, 2002). As we shall see later, $\mathcal{R}(V)$ also has a non-trivial time-structure.

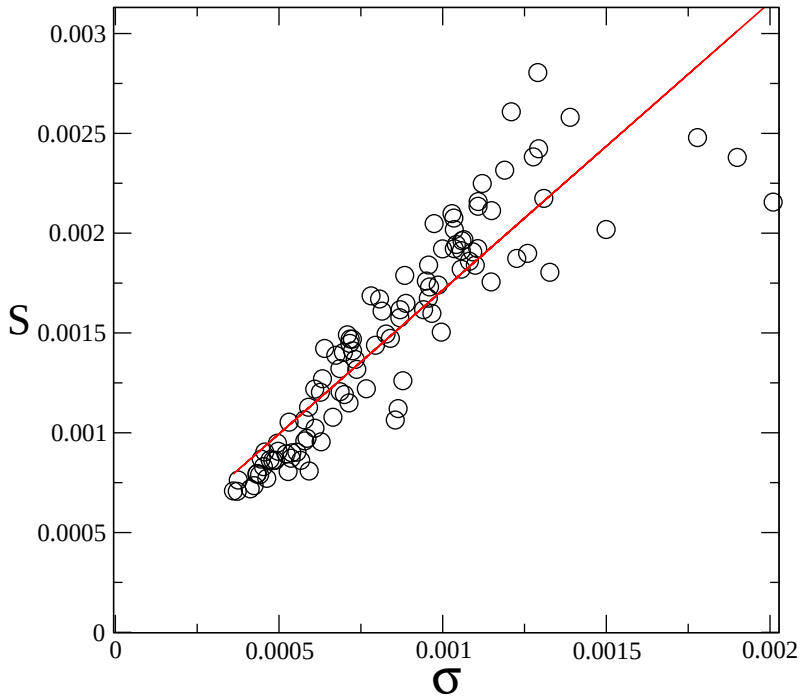

Fig. 1.1: Spread versus volatility per trade for France Telecom in 2002. Each dot corresponds to a pair $(\langle S \rangle, \langle \sigma \rangle)$ computed by averaging over 10000 trades , which is roughly two days of trading. S and σ are in relative value. The line is a fit and leads to $y = 0.00027 + 1.44 * x$. The tick size is, in this unit, 0.0006.

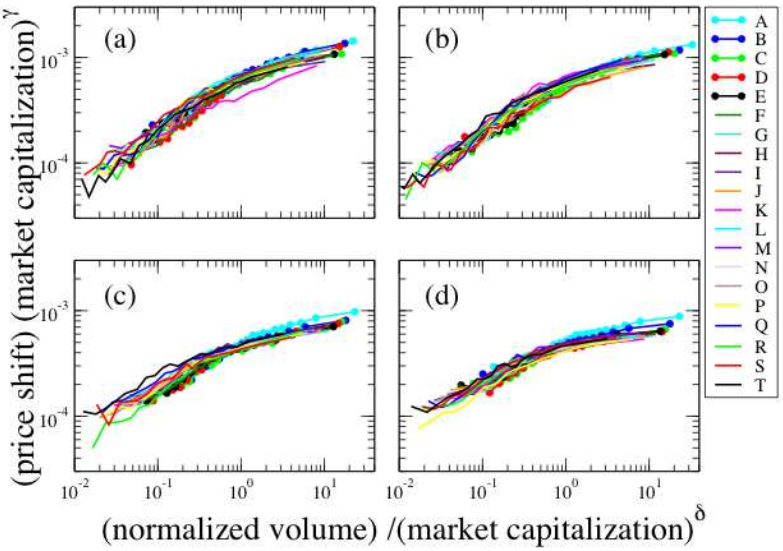

Fig. 1.2: Price shift vs. transaction size, for buy orders in 1995 (a), 1996 (b), 1997 (c) and 1998(d), renormalized by powers of the capitalization in order to make the data collapse roughly onto a single curve. Results for sell orders are very similar. Lillo et al.

Another interesting observable is the *order-book* itself. In Fig.1.3 half of the averaged order-books (the limit orders to sell) of several stocks of the Paris Bourse are presented. Using an appropriate re-scaling of the axis, this curves collapse (Bouchaud, Mézard and Potters 2002). Interestingly, the volume in the order-book varies markedly sub-linearly with the stock capitalization—maybe as a power law with an exponent 0.76 (Zumbach, 2004).

The market order flow also displays unexpected properties. As we shall detail later, the market order signs (the sign is positively defined is the market order is to buy, negative otherwise) are extremely correlated in time: the correlation decays as a power law with an exponent smaller than 1. Thus it is a *long-memory* process, in the sense that the integral of the correlation function diverges. This correlation is observable for days. By contrast with the super-diffusion that would characterize a Brownian motion with such correlated shocks, we shall explain that the correlated flow of market orders do not lead to a over-diffusion in the price fluctuations.

It is unclear if all these properties are universal and necessary laws stemming from some strategic equilibrium between agents, or if these are conventions, that may evolve in time. In the latter case, trying to derive these laws

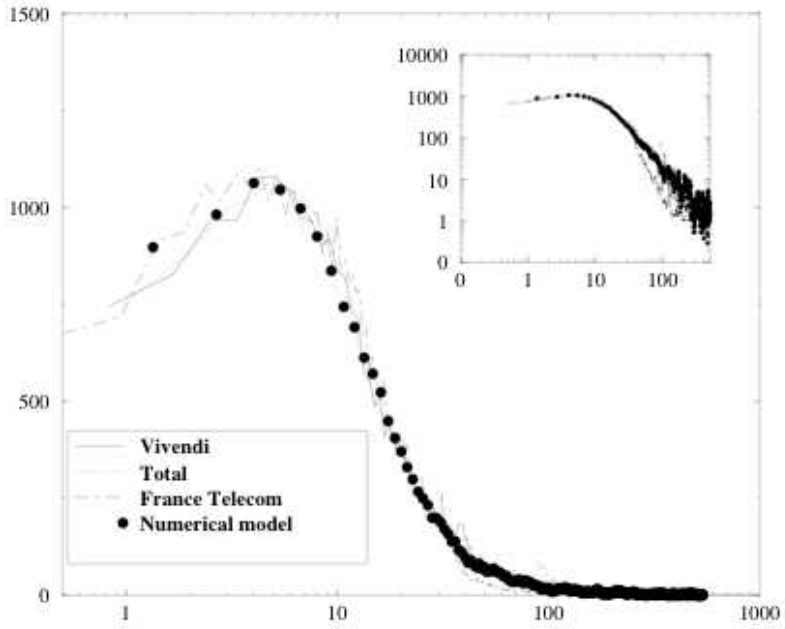

Fig. 1.3: Average volume of the queue in the order book for three stocks, as a function of the distance Δ from the current bid (or ask) in a log-linear scale. Both axis have been rescaled in order to collapse the curves corresponding to the three stocks (The thick dots correspond to a numerical model). Inset: same data in log-log coordinates (Bouchaud, Mézard, Potters 2002).

is hopeless. Then a respectable goal would be to establish necessary relations between the quantities that characterize the electronic market; which would restrict the possible micro-structural market properties. In what follows we aim to show that requiring the absence of simple statistical arbitrage enables one to understand some of these.

Microstructure has been studied with various perspectives, and by scientists of diverse origins. Economists proposed models to account for the existence and the value of the spread S (Biais, Foucault, Hillion, 1997). In these models the spread is in particular fixed by the risk aversion of liquidity providers, the "inventory control" explanation, and/or the quantity of information contained in the market orders, the "adverse selection" explanation. The drawback of these approaches is that the information content of trades, as the risk aversion of liquidity providers, are abstract notions which cannot be measured directly. Recently artificial intelligence and computer science researchers got interested in electronic markets: see for example the Penn-Lehman Trading Project (Kearns and Ortiz, 2003). This uses a software simulator for automated stock trading, the Penn Exchange Simulator, which merges automated clients orders with real-world, real-time orders. This software enables to test various strategies during open competitions (Kearns and Ortiz, 2003; Sherstov and Stone, 2004). This is thus a powerful tool to study the dynamics generated by an ecology of interacting (and possibly learning) strategies. In the future this might bring an interesting point of view to understand what determines market liquidity. In particular it would be instructive to determine what minimal set of strategies lead to a realistic market behavior. Finally, in the last few years physicists and economists have proposed to study, as a starting point, zero-intelligence models were liquidity providers and liquidity takers are mimicked by random flows, see e.g (Mendelson, 1982; Maslov, 2000; Bouchaud, Mézard, Potters 2002; Farmer et al., 2004). As we shall see, these models lead to non-trivial dynamics. Nevertheless we shall argue that they are too naive to capture the important properties of the electronic market.

Our work on microstructure is presented in Chapter 3. It contains three distinct parts:

- After describing the continuous double-auctions, we study theoretically some "zero intelligence" models previously proposed. These models were mostly studied in simulations, because they are hard to track analytically. They lead to a non-trivial price dynamics, volatility and spread. In section 2 we furnish simple qualitative arguments to estimate these quantities, and we propose a new and more realist model. We show that in zero-intelligence models the very price definition is fun-

damental. We study this question empirically. Finally, we shall argue that the zero-intelligence approach is not appropriate to describe price formation. In particular it misses the fact that imposing the absence of simple statistical arbitrage strategies constrain much the order-book properties.

- In section 3 we show that the market displays long memory: the correlation of the market order signs decay in time as a power-law with an exponent smaller than 1. Such correlation extends over *days*. This implies that the diffusive nature of price is extremely subtle: if the impact of each trade was permanent, the price would be extremely super-diffusive. We show that, since it is not the case, the impact of an *isolated* single trade must be almost completely transient, and decays slowly in time. We propose an explanation for such behavior, and we supports this claim with empirical results on the time structure of the price impact function and price diffusion. We show how the average price impact and the volatility are related — an obvious result only when the market order flow is not correlated in time.

- In section 4, we study how liquidity and volatility are related. Our main result is to show that, without assuming any underlying model on agents risk aversion, nor on the information content of trades, it is possible to relate formally spread and the average price impact. Our single assumption is that the market is competitive, and that the gain of market makers is zero. This is equivalent to impose that market orders and limit orders have on average an equal cost. We expect that it is the case in a continuous double-auctions set-up where all agents can emit both types of order. We check empirically this relation, and we find a very good agreement with our predictions.

  Since, as we show in Section 3, the impact function and the volatility are related, we obtain an equation between spread and volatility. We argue that this equation leads to a linear dependence between spread and volatility per trade, as observed empirically (see Fig1.1). This explains how volatility, liquidity and trading activity are related, which are long-standing questions of the microstructure literature.

2. Effect of self-referential behavior on price dynamics

2.1 Introduction

2.1.1 Aim of the chapter

The Efficient Market Hypothesis (EMH) posits that prices contain all available information at a given instant of time (for more precise statements, see Fama 1970). The argument invoked to support this claim is arbitrage: if prices differed from their informationally efficient value, an arbitrageur possessing some information not reflected in the price would be able to make a profit, and doing so would bring the price closer to its true value. In this framework, price changes can only be triggered by some new, unpredictable piece of information. Therefore, as shown by Samuelson, *properly anticipated prices fluctuate randomly* (prices should follow a random walk). This theory relies on the assumption that the influence of rational agents, who seek to discover the 'true', fundamental value of a stock, outweighs that of noise traders and anchors the price to its rational value. Market efficiency is of paramount importance for policy making and strategic corporate decisions and, of course, for investing in stock markets.

However, the assumptions underlying EMH are in fact rather dubious on many accounts and have been fiercely criticized by many authors in the last two decades (see, e.g. Shiller 2000 & 2002, Shleifer 2000). [1] It seems clear that agents do not act as postulated by the proponents of EMH. Apart from psychological and cognitive biases (Kahneman & Tversky 1986, Shiller 2000), herding effects (Banerjee 1992, Bikhchandani & al. 1992, Kirman 1993, Cont & Bouchaud 2000) and risk aversion that limits arbitrage (Schleifer 2000), a crucial reason, in our opinion, is that the number of objective factors that can affect the value of a stock is extremely large. The interpretation to be given to some piece of 'information' is often very ambiguous, even for agents

[1] for a recent interesting forum on this debate, see: http:// webreprints.djreprints.com /1092121283858.html and http:// www.westga.edu/ bquest /2002/market.htm.

doing their best to be rational. For example, expert financial analysts themselves are known to fare rather badly at forecasting the next earning of firms (Trueman 1994, Guedj & Bouchaud 2005). With such fragile, "gossamer" information to rely upon, market participants are more interested in guessing the opinion of the market than discovering the fundamental value of the stock. This was first emphasized by Keynes 1936 (see also Orléan 1999 for more recent work), and illustrated by his famous beauty contest, where the goal is to anticipate correctly what other participants themselves anticipate. This self-referential behavior can lead to markets that differ strongly from the predictions of EMH. An interesting example is provided by a simple game that encapsulates the basic message of Keynes' beauty concept. In this game, participants must each choose a number between 0 and 100, and the winner(s) are those whose choice is closest to one-half of the average choice (Nagel 1995). Of course, the fully rational choice is that all players choose 0. On the other hand, if agents are all totally irrational, the optimal choice is $25 = 50/2$, but if a fraction f_1 follows this first level reasoning, the optimal choice becomes $[25f_1 + 50(1 - f_1)]/2$, and so on. Empirical studies show that the average is close to 25 and that 30% of the players predict a number close to 12.5.

In the context of repeated games (such as financial markets), a natural strategy is to study empirically the statistical behavior of the other agents and to play accordingly. (Agents doing this will be called 'chartist' in the following discussion.) A common temptation is to compare the present situation with similar situations from the past and posit that what already happened is more likely to happen again. As Brian Arthur puts it: "*As the situation is replayed regularly, we look for patterns, and we use these to construct temporary expectational models or hypotheses to work with*" (Arthur 1995). As an anecdotal example, prices tend to decline before a war is declared and to rise again once the war has actually started (as the 2003 events again sadly confirmed). Often some plausible story is given to understand why such a pattern should exist. This convinces more participants that the effect is real, and their resulting behavior reinforces (or even creates) the effect (Woodford 1990): this is a self-fulfilling prophecy. A large consensus among economic agents about the *correlations* between a piece of information and the market reaction can be enough to establish these correlations. Such a 'condensation' of opinions leads to what Keynes and Orléan call a *convention*, a common lore on which uncertain agents can rely on and that supplements gossamer information. A convention may concern the overall mood of the market (bullish or bearish, for example), but it may also concern the way a piece of information is interpreted by the market. We will primarily focus on this second type of convention. The information we will

consider can either be exogenous to the market (such as the interest rate, inflation and other macro-economic figures, or geopolitical issues), or endogenous to the market (such as price patterns that feedback on the price itself including trends leading to bubbles, or ARCH-like volatility feedback, etc.)

A striking feature is that not only can these conventions spontaneously appear, but they can also disappear or even invert the purported correlation. For example, as we document in Section V below (see in particular Fig. 5), the correlation between bond markets and stock markets was positive in the past (because low long term interest rates should favor stocks), but has recently quite suddenly become negative as a new 'Flight To Quality' convention set in: selling risky stocks and buying safe bonds has recently been the dominant pattern.

The aim of the present work is to analyse a parsimonious model for the appearance and dynamics of conventions and their consequence on the statistics of price changes. As is now well known, price returns exhibit several statistical features that cannot be related to the fluctuations of any fundamental value, as should be if markets were efficient (Fama 1970). One of the biggest puzzle of EMH is the so-called 'excess volatility': Shiller (1981) showed in a famous study that the actual volatility of markets is far too large (by at least a factor 5) compared to what is expected within the efficient market theory from the volatility of dividends (see also Shiller 2000). Many (sometimes ad-hoc) attempts to reconcile this finding with EMH have been proposed, for example by adding a time dependent risk aversion factor, risk free rate, or consumption habits (Cochrane 1991, Cochrane & Campbell 1999, Black & Fraser 2000) that can be tuned to account for the observed level of volatility. However, the puzzle is not only that the market volatility level is too high, but that this volatility level itself wildly fluctuates in time, exhibiting self-similar clustering and power-law temporal autocorrelations with a small exponent. The volatility is therefore a process with long term memory (i.e. with a non-integrable correlation function), with dynamics spanning at least three decades of time scales (from hours to years).[2] The EMH description would then require in addition the news process itself to reveal such a complex, multiscale, temporal structure.

This behavior, akin to velocity fluctuations in turbulent flows (Ghashgaie & al. 1996, Mandelbrot 1997, Fisher & al. 1997, Muzy & al. 2000, Lux 2001), is remarkably universal across markets and epochs, suggesting that volatility is endogeneous to market activity and speculation, rather than governed by

[2] On this point, see: Lo 1991, Ding & al. 1993, Guillaume & al. 1997, Liu & al. 1997, Potters & al. 1998, Cont 2001, Le Baron 2001, Muzy & al. 2000, Bouchaud & Potters 2004.

external, news related shocks. Other anomalous effects (from the point of view of EMH) are worth mentioning, including the excess cross-correlations, both between domestic stocks and between international markets, that cannot be explained in terms of fundamental, economic correlations (Shiller 2000, Longin & Solnik, 2001, Ang & Bekaert 2002).

Our model is an example of a self-fulfilling process; trying to extract correlations between information and price from past observations, market participants tend to create and/or reinforce them. Using the language of physics, the model has a *phase transition*: above a certain threshold in feedback strength, the market can be in two distinct states, or two conventions. We find that this mechanism naturally leads to some excess volatility and a kind of long term memory. In the case where information is endogenous, these two states correspond to trend following or contrarian conventions where the autocorrelations are either positive (trend) or negative (contrarian). The market however switches between the two conventions on a certain time scale (that can be very long) such that *on average* the autocorrelation is zero, although *locally* the autocorrelation has a well-defined sign. These phases correspond to the market folklore: markets are indeed thought by many investors to be alternatively trending or mean-reverting in strong opposition with the prediction of EMH. Besides anecdotal evidence, we provide in this Chapter empirical evidence for the existence of these long-lived excess correlations (or anticorrelations) in stock markets (see Section V, Figs. 6, 7, 8).

2.1.2 Relation with other work and organization of the Chapter

The existence of trends and anti-trends has been broadly documented in the economic literature (see e.g. De Bondt & Thaler 1985, Hong & Stein 1999, Shleifer 2000, and references therein) where it is described as overreaction and underreaction to news. In the first case, the overreaction is later compensated by a mean reversion, whereas underreaction corresponds to an anomalously slow adjustment of the price and the appearance of a trend. A well-known study of De Bondt and Thaler that we will further discuss in the conclusion shows that over-performing stocks tend to mean-revert on the scale of 5 years, and vice-versa. Several behavioral models have recently been proposed to understand these effects (see Shleifer 2000, chapters 5 & 6, Barberis & al. 1998, Hong & Stein 1999). We follow here the same goal of articulating a simple and generic model for these pricing anomalies. Although some ingredients are common to these models and ours, there are

also major differences, both in the concepts and in their technical implementation. For example, in the 'investor sentiment' model of Barberis & al., investors postulate the existence of alternating trending and mean-reverting phases that they try to identify from observation. In our framework, on the other hand, these phases dynamically appear as agents attempt to learn the statistics of price changes from past observations. This learning aspect, much emphasized in Arthur 1995, Woodford 1990, and in the context of the Minority Game (Challet & al. 2001, 2004) is actually absent from the model presented in Barberis & al. The model of Hong and Stein postulates that some momentum traders use (but only for a limited amount of time) the positive temporal correlations created by the slow diffusion of information among news-watchers. The effect of these momentum traders is to reinforce the trend and to convert an initial underreaction into overreaction. In order to observe mean reversion effects, another category of contrarian traders must be included in the model. In our model, on the other hand, trends can appear without any fundamental news.

In this respect, it is useful to mention the work on 'sunspots' (Woodford 1991) and information cascades (Banerjee 1992, Bikhchandani & al. 1992), which also describe situations where a symmetry between different possible outcomes can be broken by a small initial (random) bias, amplified by a subsequent self-referential and reinforcing decision process. In Woodford's sunspot learning model, for example, a priori irrelevant sunspot variables become increasingly important for determining the behavior of economic agents through a reinforcing mechanism. This leads to an equilibrium where these variables *de facto* become relevant. The model proposed below does share some features with sunspot models, although it differs from them in many respects. These latter models concern production and consumption in a model economy, whereas our model focuses on the dynamics of price in a financial market. More importantly, the equilibrium reached in Woodford's model is, once reached, stable over time and conventions are eternal. Our model, on the contrary, is a model of 'punctuated equilibrium' (Bak 1996): conventions are temporary, and fluctuations can always drive the system to another state. The reason for the difference lies in the finiteness of the agent's memory in our model; introducing a similar feature in Woodford's model would presumably lead to a very interesting punctuated equilibrium situation, richer than the two-state situation described below.

Finally, several families of models where trading strategies use the price past history have recently been investigated, for example, in the schematic inductive rationality models (El Farol bar model, Arthur 1995 or Minority Games, Challet & al. 2001, 2004) or in agent based models where a fraction

of agents base their trading decision on the recent behavior of the price itself.[3] The present model is interesting because the self-referential feedback is much simpler, and its consequences can be analytically investigated in full details.

The organization of the Chapter is as follows. We describe and analyze the model in full generality in Section II, featuring agents who try to use some correlations between the price and a certain information indicator that can be exogenous or endogenous. We motivate our Langevin description of the feedback dynamics and explain how non-trivial equilibria can appear when the self-referential tendency increases. We discuss the appearance of super-long time scales for regime switching (Section III). We then move to the particular case where traders use the past price changes as a source of information (Section IV) and where the above mentioned trend following or contrarian conventions (or market sentiments) appear. We then analyze some empirical data that qualitatively support the predictions of the model (Section V). Finally, some applications and extensions of the model are proposed, and our findings are contrasted with the predictions of EMH.

2.2 Set up of the model

We will call P_t the (log-)price of a certain asset at time t, and δP_t the return between t and $t+1$. Here, $\Delta t = 1$ is the elementary time step over which agents revise their strategies, which might be one day or one week, although in some case smaller time scales such as minutes could also be usefully considered. We now argue that some agents base their strategy on the observation of the temporal change of a certain index I_t, which might be a financial index or an economic indicator (for example, dividends, interest rates, inflation, confidence, unemployment, etc.), or even, as will be considered below, the price P_t itself. We will denote as δI_t the change of this indicator between t and $t+1$. Note that δI_t could in fact be a binary variable, representing a qualitative piece of information, and that the interval Δt might not be uniform, but be the time interval between news announcements.

Suppose that there exists a causal correlation between the change of I_t and that of P_t in the sense that the correlation between δI_t and δP_{t+1} is

$$E\left[\delta I_t \delta P_{t+1}\right] \equiv C. \tag{2.1}$$

(We suppose for simplicity that all correlations for larger time lags are zero.)

[3] For a selection of papers, see Kirman 1993, Palmer & al. 1994, Bouchaud & Cont 1998, Lux & Marchesi 1999, 2000, Iori 1999, Brock & Hommes 1997, Hommes 2001, Farmer 2002, LeBaron 2002, Chiarella & He 2001, Kirman & Teyssière 2002, Giardina & Bouchaud 2002.

It is then a well known result of linear filtering that the best estimate (in a quadratic sense) of δP_{t+1} knowing δI_t is given by (see e.g. Bouchaud & Potters 2004, p. 132):

$$\delta P^*_{t+1} = \frac{C}{E\left[\delta I_t^2\right]} \delta I_t. \tag{2.2}$$

Now, we consider two types of agents, those who act randomly or based on some information uncorrelated with I_t and those who try to take advantage of the possible correlations between δI_t and δP_{t+1}. Since the true, fundamental value of the correlation C is in fact not known (and might actually be zero), agents of the second type attempt to extract this value from past history, from which they try to learn the value of C. It is natural to assume that these agents give more weight to the recent past. A convenient framework is that of exponential moving averages, such that the *estimated* value of C at time t is given by

$$C_t = \frac{1-\alpha}{\alpha} \sum_{t'=-\infty}^{t-1} \alpha^{t-t'} \delta I_{t'} \delta P_{t'+1}, \tag{2.3}$$

where α sets the memory time T of the agents, as $T = 1/|\ln \alpha|$. Other memory kernels could be considered (such as a flat window, for example), with only minor quantitative consequences on the following conclusions (at least in the limit $T \gg 1$).

Equation (2.3) is equivalent to the following Markovian update of the estimated correlation:

$$C_t = \alpha C_{t-1} + (1-\alpha) \delta I_{t-1} \delta P_t. \tag{2.4}$$

We now suppose that the agents neglect the possible fluctuations of the volatility of I_t and assume $E\left[\delta I_t^2\right]$ is a constant (that we set in the following equal to unity unless stated otherwise). Relaxing this hypothesis would again lead to minor changes in the following. The expected return between t and $t+1$ is therefore $C_t \delta I_t$. We assume that chartist agents will then buy (or sell) a quantity V_t that is an odd function of this expected return,

$$V_t = \mathcal{G}\left(C_t \delta I_t\right). \tag{2.5}$$

In general one expects the demand function $\mathcal{G}$ to be linear for small arguments and to saturate for large arguments. In the context of an exponential utility function (called CARA in the literature) and Gaussian returns, the quantity to be maximized is $r - \Lambda\sigma^2$, where r is the expected return, Λ a certain risk aversion coefficient and σ^2 the variance of the return. In this case, the function $\mathcal{G}$ is found to be strictly linear. The saturation comes from both

the limited resources of the agents and their limited ability to borrow and from an increased risk aversion for tail events. Both effects tend to limit the invested quantity even if the signal is very strong, as is intuitive.

These chartist orders add to all other ones and impact the price as

$$\delta P_{t+1} = \mathcal{F}\left(\Omega_t + NV_t\right), \tag{2.6}$$

where N is the number of chartist agents that try to exploit this correlation and Ω_t the total volume of other agents, which we assume to be a random variable of zero mean and variance σ^2. (In fact, as we will discuss below, these other agents could base their decision on different, uncorrelated information sources). The *impact function* $\mathcal{F}$ describes how a given order volume affects the price and has been the subject of many recent empirical studies (Kempf & Korn 1998, Plerou & al. 2002, Lillo & al. 2003, Potters & Bouchaud 2003). Provided the elementary time step Δt is large enough, this function is linear for small arguments and bends down for larger order imbalance. Here, we will neglect higher order contributions to $\mathcal{F}$ and simply posit, as in Beja & Goldman 1980,

$$\mathcal{F}(u) = \frac{u}{\lambda}, \tag{2.7}$$

where λ is a measure of the liquidity of the asset. Non linear corrections would only change some details of the following discussion, but our main conclusions are robust against the detailed form of both $\mathcal{F}$ and $\mathcal{G}$. Therefore,

$$\delta P_{t+1} = \frac{\Omega_t}{\lambda} + \frac{N}{\lambda}\mathcal{G}\left(C_t \delta I_t\right), \tag{2.8}$$

where C_t is self-consistently expressed as in Eq. (2.3).

As will be clear below, many of the following results only rely on the general shape of $\mathcal{G}$ (i.e. linear for small arguments with a negative curvature, possibly saturating, for larger arguments). In order to keep the mathematics simple, we expand $\mathcal{G}$ for small arguments and retain only the first two terms, a procedure that is justified in the small signal limit. Since $\mathcal{G}$ is odd, its generic expansion for small arguments reads $\mathcal{G}(u) = au - bu^3 + ...$ with $a, b > 0$. We therefore obtain the following central equation:

$$\delta P_{t+1} = \frac{\Omega_t}{\lambda} + gC_t\delta I_t - hC_t^3\delta I_t^3 + O(C^5), \tag{2.9}$$

where $g \equiv Na/\lambda$, $h \equiv Nb/\lambda$. The above equations basically describe the self fulfilling process that we study in detail now. The parameter g will turn out to be crucial in the following; note that g increases with the number of chartist agents.

2.3 Analytic results: spontaneous appearance of conventions

2.3.1 A Langevin equation

In the absence of chartist agents ($g = 0$), there are no feedback effects, and we assume that the dynamics of the price is a simple random walk (but not necessarily Brownian) of volatility $\Sigma_0 = \sigma/\lambda$. The apparent correlation C_t will describe any deviation from this trivial behavior. Using Eqs. (2.3,2.9) one finds

$$C_{t+1} - C_t = \epsilon \left(-C_t + gC_t \delta I_t^2 - hC_t^3 \delta I_t^4 + \xi_t \right), \tag{2.10}$$

where we have set $1-\alpha = \epsilon$, and $\xi_t \equiv \delta I_t \Omega_t/\lambda$ is another white noise (because Ω_t is supposed to be independent of δI_t) of zero mean and variance σ^2/λ^2.

Now, we will write;

$$\delta I_t^2 = E(\delta I_t^2) + \eta_t = 1 + \eta_t, \quad \text{and} \quad \delta I_t^4 = E(\delta I_t^4) + \eta'_t = (3+\kappa) + \eta'_t, \tag{2.11}$$

where η_t and η'_t are two correlated noises of zero mean, and κ the excess kurtosis of the index fluctuations. Therefore the evolution of C_t contains a deterministic part and a random part. In the case $\epsilon \ll 1$ considered in this Chapter, where the memory time $T \approx 1/\epsilon$ becomes much larger than the elementary time step, one can neglect, in a first approximation, the influence of η_t and η'_t (but see below). In the limit $\epsilon \to 0$, the discrete time equation (2.10) is such that $C_{t+1} - C_t \to 0$. Therefore one can write $C_{t+1} - C_t \approx dC/dt$ and take the continuum time limit to construct a Langevin-Ito stochastic differential equation for $\hat{C}_t = C_t/\Sigma_0$ in rescaled time $\epsilon t = \hat{t}$:

$$d\hat{C} = -\frac{dV}{d\hat{C}} d\hat{t} + \sqrt{\epsilon} d\xi, \tag{2.12}$$

where $d\xi$ is a Brownian noise of unit variance and we have introduced a 'potential' V given by

$$V(\hat{C}) = \frac{1}{2}(1-g)\hat{C}^2 + \frac{1}{4}(3+\kappa)\hat{h}\hat{C}^4, \tag{2.13}$$

with $\hat{h} \equiv h\Sigma_0^2$. This is the so-called Landau potential that describes phase transitions (Goldenfeld 1992). For $g < 1$ this potential has an absolute minimum at $\hat{C} = 0$, whereas for $g > 1$, $\hat{C} = 0$ becomes a local maximum and two stable minima appear for $\hat{C} = \pm C^* = \pm\sqrt{(g-1)/(3+\kappa)\hat{h}}$. Note that retaining more terms in the expansion of $\mathcal{G}$ would change the detailed shape of $V(\hat{C})$, but not the above crucial qualitative feature. It is enough to know

that $\mathcal{G}(u)$ is monotone and increases slower than linearly for large arguments to insure that $V(\hat{C})$ has the shape shown in Fig. 1. From now on, we will drop the hat on $\hat{C}$.

2.3.2 The appearance of stable conventions

From the Langevin-Ito equation for C one deduces, using standard methods (Chandrasekhar 1943, Gardiner 1996), the equilibrium distribution $P(C)$ of the Boltzmann-Gibbs form:

$$P(C) = \frac{1}{Z} \exp\left(-\frac{2V(C)}{\epsilon}\right), \tag{2.14}$$

where Z is a suitable normalization. Therefore, for $g < 1$ (weak feedback), $P(C)$ is unimodal and has a maximum at $C = 0$, whereas for $g > 1$ (strong feedback), the most probable values for C are $\pm C^*$. This means that for strong feedback, a non zero correlation between the price and the indicator spontaneously appears. This correlation can be either positive or negative, corresponding to the two possible conventions. However, on average, the correlation is still zero for $g > 1$, since C randomly flips between $\pm C^*$. In order to do so, a 'barrier' ΔV has to be crossed (see Fig. 1); the average switching time τ is well known to be given, for $T\Delta V \gg 1$, by the Arrhenius law (Chandrasekhar 1943, Gardiner 1996):

$$\tau \approx \left(\frac{T}{g-1}\right) \exp\left[2T\Delta V\right], \tag{2.15}$$

with $\Delta V = (g-1)^2/4\hat{h}(3+\kappa)$ and $T = 1/\epsilon$. Because of the exponential term in the above equation, the average switching time can be much larger than the memory time T; one non-trivial consequence of a phase transition is to generate time scales that are unrelated to the natural time scale of the problem. The convention can therefore persist for very long times. This is because the random event that would invert the signal and nucleate a new convention occurs only with exponentially rare probability. Note however that the above formula is only correct when the noise $d\xi$ is Gaussian; non-Gaussian events do accelerate the crossing of the barrier (Bouchaud & Hasareesing 2000, Lenzi & al. 2001). We will see in empirical data that extreme events may indeed be a cause of abrupt convention changes.

When $g < 1$, the distribution of C is a distorted Gaussian around $C = 0$. Neglecting the non-linear term leads to

$$C_t = \epsilon \int_0^t \mathrm{d}\xi(t')\, \mathrm{e}^{\epsilon(g-1)(t-t')}, \tag{2.16}$$

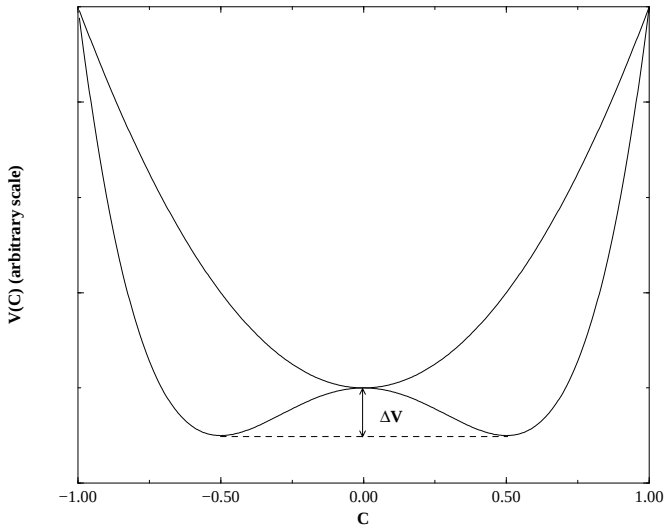

Fig. 2.1: Effective potential $V(C)$ for $g < 1$ and for $g > 1$. In the latter case, one observes two non-trivial minima at $\pm C^*$ and a 'barrier' ΔV separating them.

where we have assumed for simplicity $C_{t=0} = 0$. Hence the typical value of C (defined, e.g., as the root mean square of C) is $C \sim \sqrt{\epsilon/(1-g)}$. The typical time for the variations of C (defined from the temporal correlations of C_t) is given by $\tau = T/(1-g)$, which diverges when g gets close to 1. The chartist agents thus amplify the excursions of C, but the most probable value of C is still zero. Strictly speaking, there is no stable point or convention in this case, although when $g \to 1^-$ the excursions are of larger amplitude and of longer duration, which corresponds to what can be coined a 'floating convention'.[4]

It is interesting to give to the threshold value $g = 1$ a more intuitive interpretation. Recall that $g = Na/\lambda$, where N is the number of chartist agents and a the coefficient that relates the strength of the (apparent) signal to the investment volume. It is clear that to form a trading signal, the prediction of the future return must be compared to the volatility of the asset; therefore $a \sim v_0/\Sigma_0$, where v_0 is the average volume of investment for an individual agent. On the other hand, if the number of all other agents is N_0, one expects that the root mean square of Ω_t should vary like $\sqrt{N_0}$.

[4] We should add here a remark on the rôle of the multiplicative noise term $gC_t\eta_t$ neglected in the above analysis. Since Ω_t and η_t are uncorrelated, the Langevin noise has a variance now given by $\epsilon[1 + (2+\kappa)g^2C^2]$. Going to the Fokker-Planck equation (Chandrasekhar 1943, Gardiner 1996), one can show that for small ϵ, the rôle of this extra term is to shift the value of the critical threshold to $g = 1 + 2(2+\kappa)\epsilon$.

Therefore $\Sigma_0 \sim \sqrt{N_0} v_0/\lambda$ (assuming that all agents invest a similar volume v_0). Finally,

$$g \sim \frac{N}{\sqrt{N_0}}, \tag{2.17}$$

independently of both v_0 and λ. The conclusion is that the market enters the convention phase as soon as $N > \sqrt{N_0}$. Hence 100 correlation-hunting traders are enough to change qualitatively, or even destabilize, a market with 10000 participants.

2.3.3 Overreaction to news

Suppose now that there indeed exists a *small* objective correlation between δP_{t+1} and δI_t, justified by some real economic mechanism relating the two quantities. This means that the term Ω_{t+1}, which models the impact of news and governs the price dynamics in the absence of chartist traders, and δI_t have a non-zero objective correlation coefficient:

$$E[\frac{\Omega_{t+1}}{\lambda}\delta I_t] = \beta E[\delta I_t^2]. \tag{2.18}$$

We conform here to the common usage of calling this particular correlation coefficient the 'beta'. The effect of such a term is to add a linear contribution to the effective potential $V(C)$ of the Langevin equation that breaks the symmetry $C \to -C$:

$$V(C) \longrightarrow V(C) - \beta C. \tag{2.19}$$

For $g < 1$ and β small, the most probable value of C is $\beta/(1-g)$ ($= \beta$ for $g = 0$, as it should). Therefore, C is of the same order as its true cause whenever $g < 1$. However, in the limit $g \to 1^-$, the apparent correlation that arises becomes much larger than its true cause: the sensitivity of the market to external information is anomalously amplified. For $g > 1$, the term βC breaks the symmetry between the two conventions $\pm C^*$. In the limit $\epsilon \to 0$, the most probable value of C is given by $+C^*$ for $\beta \to 0^+$ and by $-C^*$ for $\beta \to 0^-$ (see Equation 2.14). Therefore, in the convention phase, the amplitude of the apparent correlation is totally unrelated to that of the true correlations, although the *sign* of the correlation reflects the underlying economic reality. One observes here a typical example of overreaction to news leading to excess correlations that are well documented in the literature (De Bondt & Thaler 1985). For example, the correlations between the stocks belonging to an index and the index itself are too strong to be explained by the intrinsic correlations between the stocks (Shiller 2000, p. 186-189). The first quarter of 2003 offers a good illustration of this effect: the cross correlations

between U.S. stocks was at a historical high; a possible interpretation is that due to the large uncertainty, traders' hunt for useful information was more acute, and the influence of the index on individual stocks was expected to be anomalously large. In our model, this corresponds to the case where the indicator I_t is the stock index, a case detailed in section 2.4.4. Another well known example explores the excess correlation (in particular in crisis periods) between emerging country markets belonging to different geographic regions (Longin & Solnik 2001). In order to understand the relation between these effects and the present model, we need to stress that although C_t is a lagged correlation between unequal times, the *equal time* correlation between δP and δI measured on a coarser time scale will reflect the value of the lagged correlation C. In other words, causal correlations on a fine time scale do generate equal time correlations on a coarser time scale. More precisely, one has, for the coarse-grained covariance,

$$E\left[(P_{t+n}-P_t)(I_{t+n}-I_t)\right] = E\left[\sum_{t'=t}^{t+n-1}\delta P_{t'}\sum_{t'=t}^{t+n-1}\delta I_{t'}\right] = \sum_{t'=t+1}^{t+n-2} C_{t'} \approx (n-1)C_t \tag{2.20}$$

where the last equality holds if $n\Delta t \ll \tau$ (i.e. when the coarse time increment $n\Delta t$ is small compared to the convention shift time τ). Therefore, if strong causal correlations are established intra-day, as is the case between individual stocks and the index, a normalized excess same-day correlation between stocks $\approx C_t$ will also appear; see section 2.4.4.

2.3.4 Consequences for the price fluctuations: excess volatility

The feedback effect leads to an increase of the volatility of the price, since the instantaneous volatility is given by

$$\Sigma_t^2 = E[(\delta P_t)^2] = \Sigma_0^2(1+g^2C_t^2) + O(C^4). \tag{2.21}$$

The non-trivial dynamics of C therefore leads to a volatility increase that can be substantial in the convention phase. This mechanism, interestingly, also leads to volatility fluctuations (or 'heteroskedasticity'). These volatility fluctuations are characterized by the correlation time τ, which become large when g approaches or exceeds the threshold value $g = 1$ (see Eq. (2.15)).

The above mechanism can easily be extended to the case where agents scrutinize M different sources of information, say I_t^k, with $k = 1, ...M$. If the variation of these different indices are uncorrelated, it is easy to see that the

simultaneous effect of all the different feedbacks leads to a volatility given by

$$\Sigma_t^2 = \Sigma_0^2 \left(1 + \sum_{k=1}^{M} g_k^2 C_{k,t}^2 \right) + O(C^4) \tag{2.22}$$

(with obvious notations). Therefore, the volatility can be substantially increased if a large number of information sources are overly interpreted. Within the context of efficient markets theory, all decisions are based on some real information. This corresponds, in the above formula, to $g_k \equiv 0$ and $\Sigma_0^2 = \sum_{k=1}^{M} \beta_k^2$ where β_k describes the true causal relation between the economic indicator I^k and P. What we have shown here is that because of the feedback loop, $g_k > 0$, the empirical correlations (which are the only way to learn the value of the β_ks in the absence of any firmly grounded theoretical model) are distorted and amplified, leading to a much larger apparent $C_{k,t}$, and therefore to a potentially considerable increase of the volatility as compared to its fundamental value.

We believe that the above scenario for self-referential speculation is generic. When strategies are built using the outcome of past random events, a feedback loop can appear and destabilize the market from its putatively efficient behavior. If the feedback is strong enough, a non-trivial equilibrium can set in where self-fulfilling prophecies can establish and survive. These conventions can have no rational basis whatsoever or be the result of the amplification of a very small, but indeed objective correlation.

2.4 Price based strategies and market phases

2.4.1 Motivations

As recalled in the introduction, the basic tenet of the theory of efficient market is that prices instantaneously reflect all useful information. However, since all market participants face impact and slippage issues (due to the finite liquidity of any traded asset), those who believe that they have some useful information about future price changes must use it in such a way that their very action does not perturb the market too much. Otherwise, the potential gain associated with this information cannot be realized or only using small volumes. Therefore, informed investors must, to some extent, dilute their order in time (Kyle 1985). Doing so, they create positive temporal correlations, see the next Chapter; the slow incorporation of information into price is the underreaction phenomenon described in Shleifer 2000, Hong &

Stein 1999. Other participants that see an increase of price can believe that it is due to some information not yet available to them, but that is reflected by the recent price change. These participants will be tempted to 'jump in the bandwagon' and act as trend followers; this is at the heart of the models developed in Shleifer 2000, Hong & Stein 1999. Conversely, large orders in temporarily illiquid markets might affect the price too much (overreaction), and some restoring trades will later on move the price back to more realistic values.

Hence, there might indeed be deep reasons why it is useful to watch past price changes and be influenced by them. That this is the case in practice is beyond any doubt and is confirmed by casual observation of traders in market rooms and by several formal surveys (Shiller 2000, p. 47). In fact, it seems that price itself is, for many traders, the most relevant source of information (if not the only one, in the case of some hedge funds using statistical methods). As in many previous models, we thus consider that the information used by some agents to predict future prices is the past price change itself. However, at variance with some of these models, the economic reality of the correlations is in fact not needed, since in the strong feedback phase these correlations may spontaneously appear.

2.4.2 Trend following and contrarian conventions

In this section we study the model where $\delta I_t = \delta P_t$. In this case, the above correlation coefficient C_t becomes the autocorrelation of successive price changes. The above analysis is almost unchanged, up to a renormalisation of the coefficient h that appears in the non linear term hC_t^3. This comes from the fact that the denominator in the linear filter, namely $E[\delta P_t^2]$, is now itself affected by the feedback effect. Hence, the phase transition found above for $g = 1$ is also present in this case. In the convention phase $g > 1$, the two states of the markets correspond to positive autocorrelations ($C = +C^*$), which can be called a *trend follower* phase where past price changes tend to be followed by a change of the same sign or to negative autocorrelations ($C = -C^*$) in the *contrarian* phase, where past price changes tend to be followed by a change of opposite sign. Let us emphasize that a trend following period is not necessarily a period where the price steadily increases (or decreases), but rather a period where successive price changes have a large probability to be of the same sign (see the central period in Fig. 2.2, corresponding to $C > 0$).

We show in Fig. 2.3 the histogram of C_t from the numerical simulation of Eq.(2.9) with Ω_t a white Gaussian noise and for two values of g. Note that without a symmetry breaking term, the average autocorrelation is zero

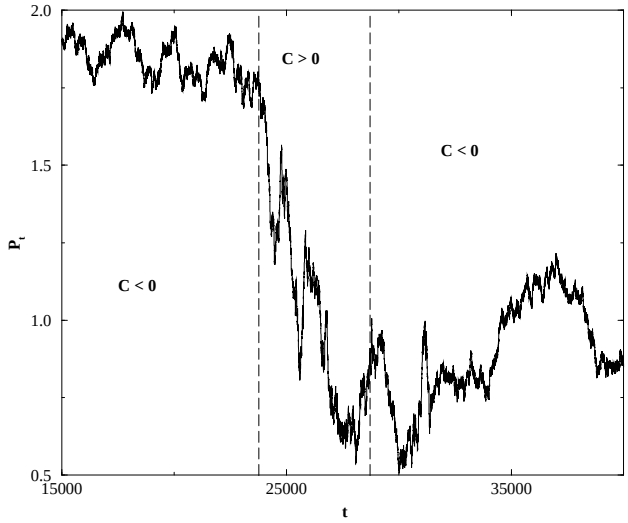

Fig. 2.2: Example of a synthetic price history, with two convention changes, for $g = 1.2$ and $\epsilon = 0.01$. Note that the coarse grained volatility is smaller in contrarian phases ($C < 0$) and larger in trend following phases ($C > 0$).

even for $g > 1$. We have here an interesting statistical process: the long time autocorrelation is zero, but local trends or anti-trends can appear and remain for quite long periods. As mentioned in the introduction, this corresponds to the market folklore: practitioners often talk about market phases where trend following strategies are supposed to be profitable and market phases where contrarian (mean reverting) strategies are supposed to work. Interestingly, any long term analysis of the average correlation coefficient would fail to reveal such phases.

2.4.3 Consequence on the volatility

An important consequence of the existence of conventions is that the coarse-grained volatility can be different from the instantaneous one. As above, the instantaneous volatility is increased compared to its bare value Σ_0 and given by $\Sigma^2 = \Sigma_0^2(1 + g^2 C_t^2)$. On the other hand, the coarse-grained volatility Σ_{cg}, defined on an intermediate time scale T^* such that $1 \ll T^* \ll \tau$ (such that C_t itself has not evolved significantly), is easily calculated. From Eq. (2.9), we deduce that

$$E[\delta P_{t+n} \delta P_t] = (g C_t)^n \Sigma^2; \tag{2.23}$$

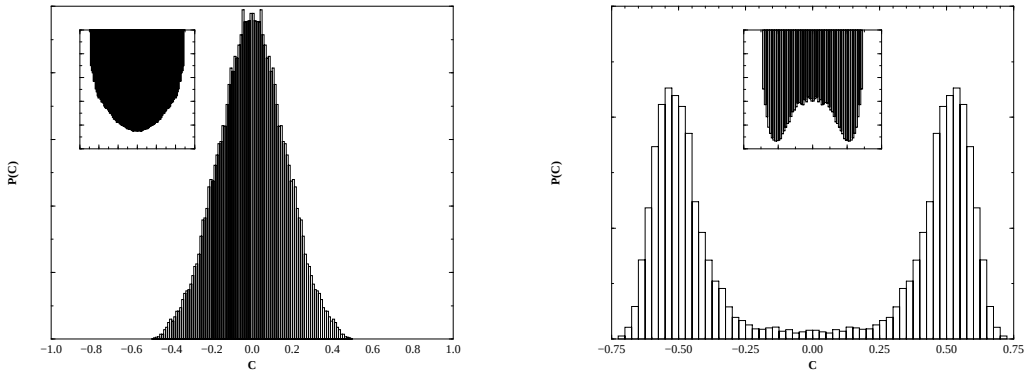

Fig. 2.3: Left: Correlation histogram $P(C)$ for $g = 0.9 < 1$ and $\epsilon = 0.01$. Right: Correlation histogram $P(C)$ for $g = 1.3 > 1$ and $\epsilon = 0.01$. Insets: Effective potential $V(C) = -\ln P(C)$.

therefore for large N,

$$N\Sigma_{cg,t}^2 \equiv E[(\sum_{i=1}^{N} \delta P_i)^2] = \Sigma^2 \left(N + 2 \sum_{i=1}^{N-1} (N-i)(gC_t)^i \right) \approx N\Sigma^2 \frac{1+gC_t}{1-gC_t}. \tag{2.24}$$

Therefore,

$$\Sigma_{cg,t}^2 \approx \Sigma^2 \frac{1+gC_t}{1-gC_t}. \tag{2.25}$$

This shows that the volatility is increased in the trend following convention and decreased in the contrarian convention. This is illustrated in Fig. 2.2, where we show the result of a simulation corresponding to $g = 1.2$ and $\epsilon = 0.01$, with a Gaussian noise term Ω_t. Note that the true long time square volatility is equal to the time average of $\Sigma_{cg,t}$ and is dominated by the trend following phases.

Equation (2.25) shows that the volatility can have large fluctuations and long term correlations; in particular, in the $g > 1$ phase, there are two time scales that govern the evolution of C_t. A relatively short one governs the fluctuations of C_t around the dominant convention $\pm C^*$; the other that can be much longer is given by the flip time τ between the two conventions, Eq. (2.15). It might be tempting to relate this to the well known fact that empirical volatility fluctuations reveal non-exponential, multiscale relaxation in time, although other interpretations exist, see e.g. (Giardina & Bouchaud 2002).

Suppose that one is in the trend following convention. The average duration $\mathcal{T}$ of a trend can be obtained by comparing the value of the coarse grained volatility to the instantaneous one. A random walk with a persis-

tence length $\mathcal{T}$ and local volatility Σ has a coarse-grained volatility given by $\Sigma^2_{cg,t} = \mathcal{T}\Sigma^2$. Using the above result, this leads to

$$\mathcal{T} \approx \frac{1 + gC^*}{1 - gC^*}. \tag{2.26}$$

Hence, one can observe two types of dynamics when $g > 1$. If the change of convention is faster than the typical duration of a trend (i.e. if $\tau \ll \mathcal{T}$) one obtains the dynamics shown in Figure **??**-b, where a period of low volatility is followed by a few sudden trends, which can be of either sign. Note that this can only occur if ϵ is large enough, which corresponds to a very short memory time (in other words that agents over-focus on very recent events). The phenomenology is in this case quite different from that shown in Fig. 2.2, where the price displays many trends before changing conventions.

2.4.4 A special case: regressing on the index

It is interesting to discuss the special case where the information I_t is the stock index itself. It is clear that in practice, the evolution of stock prices on short time scales is strongly affected by the index, which is an immediately available piece of relevant information for all market participants. Let us call P^j_t $(j = 1, ..., M)$ the price of the j-th stock belonging to the index. Then, assuming the index is computed as a equi-weight average over all the stocks, one has

$$\delta I_t \equiv \frac{1}{M} \sum_{j=1}^{M} \delta P^j_t. \tag{2.27}$$

On the other hand, the feedback effect of the index on the stock price can be written as

$$\delta P^j_{t+1} = \tilde{\Omega}^j_t + g_j C_{j,t} \frac{\delta I_t}{\Sigma^2_I} + O(C^3_j), \tag{2.28}$$

where $\tilde{\Omega}^j_t = \Omega^j_t/\lambda$ results from the trading not based on the index, $C_{j,t}$ is the empirical covariance between δP^j_{t+1} and δI_t, and Σ_I is the index volatility. Using Equation (2.27) one therefore finds, in the simple case where all g_j are equal,

$$\delta I_{t+1} = \frac{1}{M} \sum_{j=1}^{M} \tilde{\Omega}^j_t + gC_t \frac{\delta I_t}{\Sigma^2_I}, \tag{2.29}$$

where $C_t = \sum_j C_{j,t}/M$ is the covariance between δI_{t+1} and δI_t.

In the case of the index, the analysis of tick by tick data shows that the feedback takes place at very high frequencies and $\Delta t = 1$ corresponds here to

a few minutes. Summing (2.27) from t to $t+n$ defines the return on a coarse-grained scale ΔI_t. Assuming that $n\Delta t \ll \tau$, such that C_t is approximately constant, leads to

$$\Delta I_t = \frac{1}{M}\sum_{j=1}^{M} \hat{\Omega}_t^j + g\hat{C}_t \Delta I_t, \tag{2.30}$$

which is valid when $n \gg 1$. In the above equation, $\hat{\Omega}^j$ is the aggregate of the noise $\tilde{\Omega}^j$ over the time interval $[t, t+n]$, and $\hat{C}_t = C_t/\Sigma_I^2$. Finally,

$$\Delta I_t = \frac{1}{M(1-g\hat{C}_t)}\sum_{j=1}^{M} \hat{\Omega}_t^j. \tag{2.31}$$

If the Ω_t^j were uncorrelated from one stock to the next and in the absence of feedback, the index volatility would be very small compared to that of stocks (of order $1/\sqrt{M}$). Empirically, though, the S&P 500 index volatility is found to be as high as a third of the individual stock volatility. Of course, one expects that the Ω_t^j are somewhat correlated, reflecting a common sensitivity to news. However, as noted in Shiller 2000, the correlation between stocks expected from fundamental analysis is insufficient to explain the observed correlation (and therefore the volatility of the index). The model presented here shows that a high frequency positive feedback leads to an increase of the index volatility by a factor $1/(1-g\hat{C}_t)$, which can be large. This increase is actually larger than the increase of the volatility of individual stocks induced by the above feedback (the factor is in that case only $\sqrt{1+g^2\hat{C}_t^2}$).

2.4.5 Statistical Arbitrage?

As the chartists create correlations in the price, one may ask if other agents could use these correlations to make statistical arbitrages. In practice, such arbitrages are possible only if (i) the signal is strong enough in comparison with transactional costs and (ii) the time scale where correlations exist is large (because transactional costs increase at high-frequency). If the information is exogenous and continuous, as is the bond price for the stock market, we expect correlations to be established quasi-instantaneously (that is, on the time scale of the minute), possibly without transaction. In this case, even if the signal is strong, see for example Fig.2.4, no consequent statistical arbitrage can occur. As we discuss in the present section the information can also be endogenous. In Fig.2.5 we show the daily correlations of the Dow-Jones during the last century. Since the signal is daily, statistical arbitrage is possible if the signal is strong. It was the case before 1985. A possible explanation is that transactional costs were much larger at that time, and quantitative

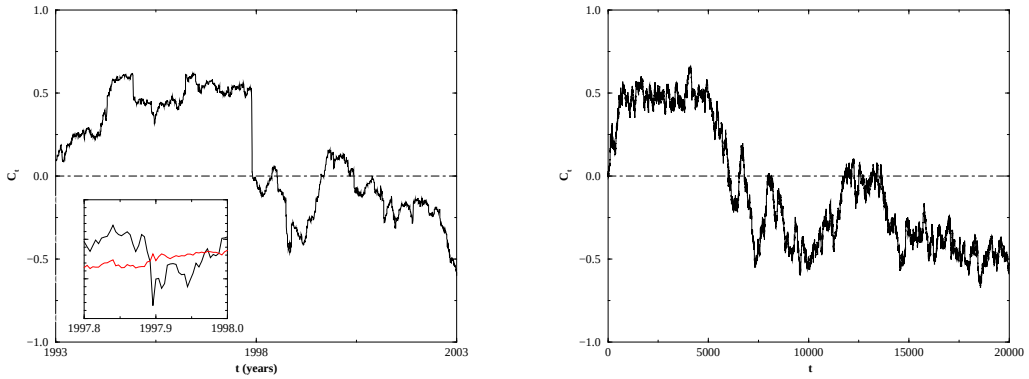

Fig. 2.4: Left: Normalized correlation between the Dow-Jones daily returns and the daily returns of a U.S. bond index with 7 to 10 years bonds, computed with $\epsilon = 0.01$. Note the convention change occurring at the end of 1997. Inset: Evolution of the Dow-Jones and the bond index in the last quarter of 1997. Right: Time dependent correlation C_t in our model, for $g = 1.2$ and $\epsilon = 0.01$.

data less easily available. Nowadays, transactional costs are much smaller, and this signal almost disappeared *on average*.

2.5 Empirical evidence

The aim of this section is present some empirical data that qualitatively support our contention that some anomalous correlations exist in financial markets with persistence times which can be very long (ten years or so). We first present the case of the bond index vs. stock index correlation, which is interesting from the point of view of the present model because it might represent an empirical realization of the convention shift scenario predicted above. We then turn to the analysis of the (daily) lagged autocorrelations of the Dow Jones index during the 20th century, which are clearly found to be significant and time dependent (positive – trend following, or negative – contrarian). Finally, we turn to single stocks cross-correlations and also find evidence for different regimes.

2.5.1 The bond/stock cross correlation

A very interesting example of rapid convention change has taken place in the 90s and concerns the correlation between stock markets and bond markets. The usual argument is that as long term rates fall, not only does holding

bonds becomes less profitable (bond prices rise) but also borrowing long term money becomes cheaper. Therefore stock markets become more attractive, and stock prices rise; this leads to a *positive* correlation between bond price changes and stock price changes. We compute the time dependent autocorrelation C_t as an exponential moving average, as given in Eq. (2.3), where I_t is the bond index and P_t is the log price of the Dow-Jones. This correlation is indeed found to be positive and very strong (≈ 0.5), in the beginning of the nineties (see Fig. 2.4). However, another story now seems to be dominant: a fall in stock markets signals an increased anxiety of the operators who sell their risky paper and buy non-risky Government bonds. This has been called Flight to Quality (FTQ). The result is a *negative* correlation between stock prices and bond prices. Fig. 2.4 shows very clearly that a change of convention has taken place in late 1997; the negative correlation was even stronger in early 2003. Quite interestingly, this convention shift has taken place very abruptly due to a series of extreme events both on the stock market and on the bond market (see the inset of Fig 2.4), in qualitative agreement with the model discussed in this Chapter. [Note that we consider here equal time correlations of daily returns; using the argument presented in section 2.3.3, we expect a high frequency causal correlation to manifest itself as an equal time correlation on a coarser time scale.]

2.5.2 The Dow Jones

We considered the detrended Dow-Jones index in the period 1900-2003, when the average return was subtracted. We have actually first fitted the log Dow-Jones as a second order polynomial in time, since the average return itself seems to have significantly increased between 1900 and 2000. [5] We again compute the time dependent autocorrelation C_t as an exponential moving average, as given in Eq. (2.3), with now $\delta I_t = \delta P_t$, and where P_t is the log price. Since the returns are non-Gaussian, we compared all our results with a null hypothesis benchmark where all returns are multiplied by random independent signs, such as to keep the correct statistics of the amplitudes but remove all serial correlations. (Note however that in this procedure, the strong serial correlations in the volatility *are* preserved.)

We show in Fig. 2.5 the time evolution of C_t computed for $\epsilon = 0.001$ for the real time series. One clearly sees that (i) C_t can be substantially larger than expected if no correlations were present and (ii) the time scale for the evolution of C_t can be much larger than $1/\epsilon \approx 3$ years. Plateaus that

[5] Taking the raw returns without detrending in fact leads to the same conclusions. The reason is that the typical daily returns are much larger than the average trend anyway.

last several decades can be observed. The histogram of different values of C_t is shown in Fig. 2.4 and is markedly different from the one corresponding scrambled series, for which all correlations are killed. (For example, the hypothesis that the two distributions are the same is strongly rejected by the Kolmogorov-Smirnov test). The century was dominated by a positive correlation convention, especially between the 50s and the 80s. Nevertheless the negative correlation convention seems to appear after 1929 during the Great Depression. There are also regimes where C_t is close to zero. This suggests that in fact g has varied over the years, with periods where $g < 1$, with no clear trends nor anti-trends appearing, and periods where $g > 1$, during which the market is temporarily locked in one convention or the other. In order to check whether the plateau values appearing in Fig. 2.4 do indeed correspond to conventions, we have determined the probability distribution of C_t with a smaller averaging time of 100 days ($\epsilon = 0.01$), and in restricted periods of time: (a) at the beginning of the 30s (contrarian convention) and (b) between the 1950 and 1980 (trend following convention); see Fig. 2.6. The comparison with scrambled data indeed shows a clear assymetry in both cases that should not exist if all serial correlations were zero.

These curves show that conventions can persist up to 30 years. The change in convention can be rather smooth, as during the second part of the century. As we saw before, the value of the most probable value C^* is related to g, that is to the number of agents using a self-referential strategy. Then these smooth changes may be due to a continuous change in the number of these agents. A change of convention can also occur suddenly, triggered by an extreme event, as it did after 1929. It can be explained as suggested above: before the crash, g is smaller than unity, and no clear convention exists. The crash induces an enormous uncertainty about the true value of stocks and encourages agents to pay more attention to past price variations. This may have led to a substantial increase of g that favored the appearance of a contrarian convention.

From the data, it appears that there might be a systematic bias towards the trend following convention. One in fact expects that some symmetry breaking term, favoring $C > 0$, should exist in general. First, as mentioned in section 2.4.1, there might be good reasons to think that positive correlations are indeed created by the time dilution of large orders or other mechanisms. Also, for purely psychological reasons, trend following strategies are more likely to be learnt and adopted than contrarian strategies because the pattern is much more obvious. This can be modeled by postulating that g depends on the sign of C_t, with $g_+ > g_-$. Therefore, nicely symmetric histograms such as those presented in Fig. 2.3 are unlikely to be observed in real markets.

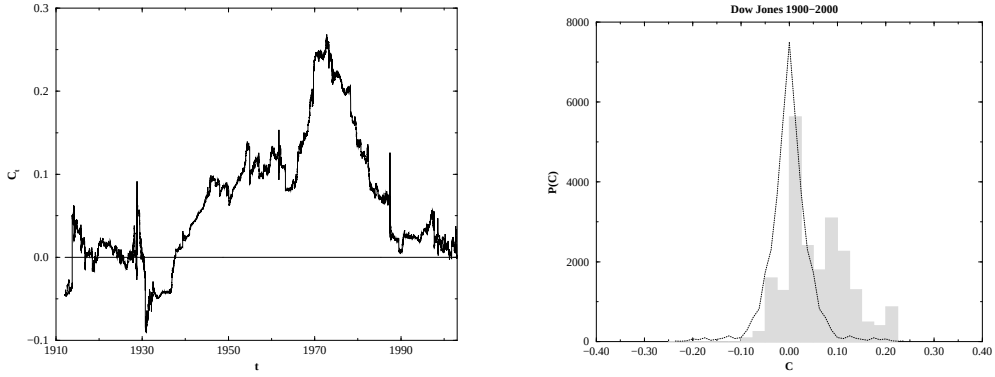

Fig. 2.5: Left: Historical time series of the daily autocorrelation C_t of the Dow-Jones index, computed with $\epsilon = 0.001$. Right: Correlation histogram $P(C)$ for the Dow-Jones with $\epsilon = 0.001$, compared to the histogram computed with the same data and the same value of ϵ but with returns multiplied by random signs (dotted line).

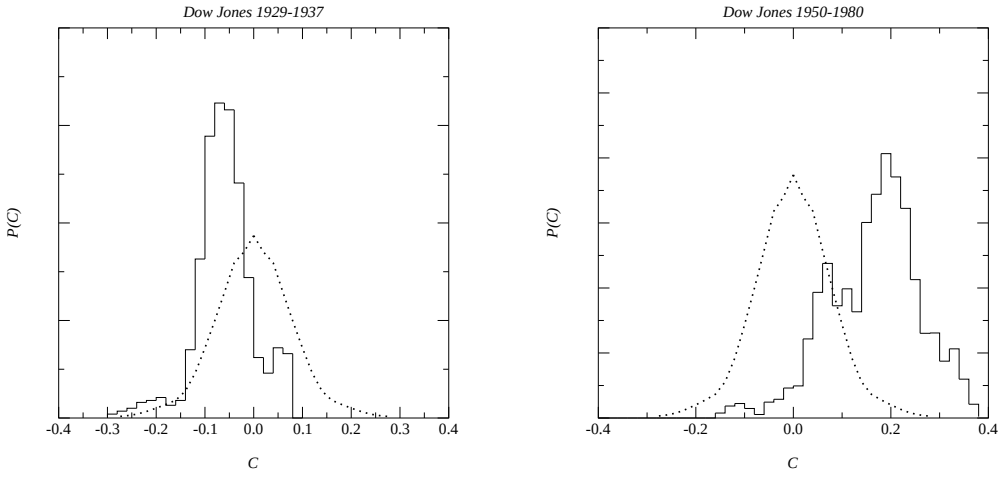

Fig. 2.6: Left: Correlation histogram $P(C)$ for the Dow-Jones in the post-crash period 1929-1937, with $\epsilon = 0.01$, compared to the 'zero correlation' histogram computed by multiplying the returns by random signs (dotted line). Right: Correlation histogram $P(C)$ for the Dow-Jones in the trend following period 1950-1980, with $\epsilon = 0.01$, again compared to 'zero correlation' histogram (dotted line).

2.5.3 Stock cross-correlations

Finally, we have studied the cross-correlations between pairs of stocks, or between a stock and the index, for the S&P 500 stocks in the period 1996-2003. We have studied daily returns and chosen an averaging time of 20 days ($\epsilon = 0.05$). Since different pairs of stocks have different average correlations, we have studied a centred, rescaled correlation, defined as

$$\tilde{C}_t = \frac{C_t - \overline{C}}{\sigma_C}, \tag{2.32}$$

where $\overline{C}$ is the average (over time) correlation between the two stocks under consideration (or a stock and the index), and σ_C the root-mean-square of this correlation. We then pool together all stocks to generate a single histogram of the values of $\tilde{C}_t$. We again generate, for comparison, surrogate data sets where the average correlation between stocks is kept, but where the residuals have random signs. The result is that the empirical distribution of $\tilde{C}_t$ is markedly different from the surrogate data; see Fig. 2.7. The empirical distribution has a clear negative skewness, whereas the surrogate data is very close to Gaussian. Restricting to highly correlated stocks ($\overline{C} > 0.3$, $\sigma_C < 0.2$), the distribution is seen to acquire a double hump structure (see Fig. 2.7, inset) that suggests the coexistence of periods where stocks are more correlated with periods during which this correlation is less. We have checked that many different pairs (not all containing the same stock) contribute to the secondary hump. This effect is probably due to a sectorial change of correlation.

2.5.4 Discussion

The conclusion of this section is that several financial market stylized facts do indeed support, in a qualitative way, the predictions of our model, although of course other explanations may exist. It would be interesting to devise more rigorous statistical tests, although we know that these usually require strong assumptions that we do not necessarily wish to make. For example, as we have emphasized above, there are many reasons to believe that the very parameters of the model (in particular g) should be time dependent. Still, with the possibility of a systematic analysis of the impact of news (economic figures) on the dynamics of financial markets, in the future one might be able to test more critically our statistical convention model.

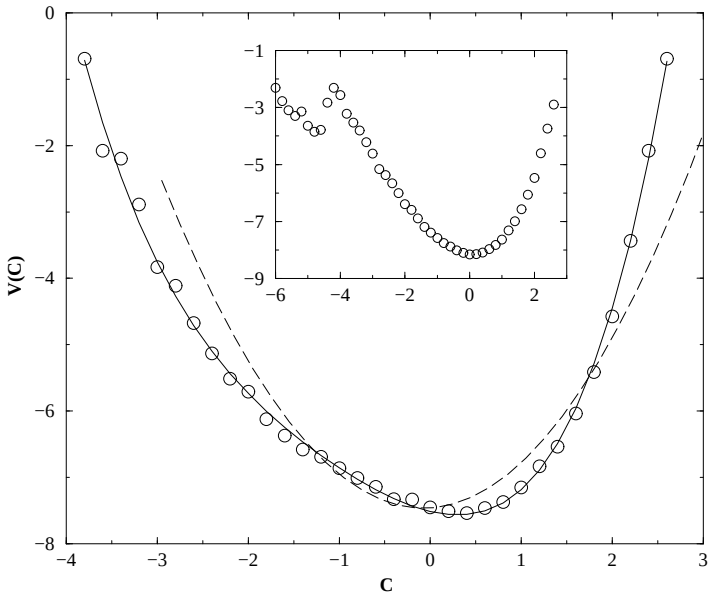

Fig. 2.7: Main figure: log-Histogram $-\ln[P(\tilde{C})]$ of the centred and rescaled correlation between stocks of the S&P 500 and the index, during the period 1996-2003, measured with $\epsilon = 0.05$. The dotted line corresponds to 'zero correlation' surrogate data. The plain line is a fit with a biaised quartic potential $V(C) = -\beta C + \alpha C^2 + \gamma C^4$. Inset: log-Histogram of the centred and rescaled correlation $\tilde{C}$ between pairs of highly correlated stocks ($\overline{C} > 0.3$, $\sigma_C < 0.2$), suggesting a bimodal structure.

2.6 Conclusion and Perspectives

In this Chapter, we have defined and studied a generic, parsimonious model that describes the feedback effect of self-referential behavior. If sufficiently strong, this feedback destabilizes the market, and non-trivial correlations can spontaneously appear or be anomalously amplified. In this case the market enters a new equilibrium state where strong correlations between a priori uncorrelated quantities might self-consistently establish. These anomalous correlations lead to both excess volatility (that may display long memory) and excess cross correlations. An interesting outcome of our model is (i) the existence of correlations with an amplitude unrelated with any rational value, and (ii) the appearance of a very long switching time scale, unrelated with the natural time scales of the dynamics (i.e. the decision time scale Δt or the memory time T). Therefore, our model displays *regime switching* over long times scales, which by the way justifies why agents should use a finite memory time in order to measure correlations in an ever changing environment. It is also worthwhile emphasizing that if the price itself is used as a source of information, some linear autocorrelations do appear on intermediate time scales, but average out on time scales larger than this switching time. In other words, the price process has zero average linear correlations, but non-zero *local* autocorrelations (trending or mean reverting). We have presented rather convincing empirical evidence that such conventions may indeed exist in financial markets. One of the most compelling cases concerns the correlation between stock markets and bond markets where both market states can be observed; the correlation appears to have rapidly shifted in the last decade from being strongly positive to being negative. An interesting application of the model could be to understand quantitatively the dynamics of inter-stock correlations, which is of crucial importance for asset allocation, risk control, and index option pricing.

Our theoretical model could be extended in several directions. First, one could consider serial correlations beyond the elementary time lag Δt, say between increments lagged by $n\Delta t$. At the linear level, the stability criteria is easily shown to be $g_n < 1$, where g_n is the feedback strength corresponding to lag $n\Delta t$. However, interesting non-linear effects can appear. For example, with two lags $n = 1$ and $n = 2$, one can observe a 'first order' phase transition where the most probable value of C discontinuously jumps from $C = 0$ to $\pm C^*$, with $C^* > 0$ even close to the transition. This is distinct from the 'second order' scenario explored in the present Chapter, where $C^* \sim \sqrt{g-1}$. Second, the case where the different sources of informations δI_t^k are themselves asset prices is quite interesting since the feedback loop also affects the cross-correlation between the different assets. Non-trivial coupled

convention dynamics can set in, in particular when the number of assets is small.

An ingredient that should be implemented is the feedback between the values of the coupling parameter g and memory time $1/\epsilon$, and the past price dynamics itself. As we have seen above, the value of g is related to the number of agents (or more precisely the total volume of orders) that act in a self-referential way. It is clear that both in periods of large uncertainty (after a crash, for example) or within a speculative bubble where the trend following strategy appears to be successful, one expects the value of g_t to grow. (A similar mechanism was recently considered within the 'Grand Canonical' version of the Minority Game; see Jefferies & al. 2000, Bouchaud & al. 2001, Challet & Marsili 2003). It would be interesting to study a precise model where the dynamics of g_t and that of the price and volatility are explicitly coupled, in the spirit of Bouchaud & Cont 1998. One can expect that such a model would be able to capture a lot of the financial markets phenomenology. Along similar lines, if there are several sources of information δI_t^k, one should expect that the feedback of successful strategies onto the value of the couplings $g_{k,t}$ will be unstable in the sense that one of the g_k will grow at the expense of the others because the coordination of strategies leads to stronger self-fulfilling prophecies and therefore larger potential profits. In other words, this feedback between the tendency to follow a pattern and its predictability leads to a condensation of the strategies in a few prominent conventions, with abrupt transitions between those.

Finally, we need to discuss the above model from the point of view of Efficient Markets and show how it should be modified to describe the *long term* behavior of market prices. If the information is systematically overinterpreted and the volatility much too large compared to that of the fundamental value, the price should go on long time scales to completely unrealistic values. More precisely, the difference between the fair price and the market price would typically grow as $\sqrt{T}$. (In mathematical terms, the market price and the fundamental price would not cointegrate.) The answer to that paradox is, we believe, the following: nobody knows the fair price of a stock more accurately than within, say, a factor of two or three. This was actually proposed by Black 1986 (somewhat humoristically) as an operational definition of an efficient market, and this view seems to us to be fundamentally correct. For example, the historical analysis presented in Shiller 2000 (p. 8), shows that the price to earning ratio of U.S. stocks has indeed fluctuated, from 1900 to 2000, between 10 and 40, so it is reasonable to think that there is a wide band of prices across which arbitrage cannot take place because of the lack of a reliable estimate of what the true price should be. As emphasized by Shleifer and others, arbitrage only makes sense if one can compare

the *relative* price of two assets, but becomes very dodgy if one speaks about absolute values. Therefore, one expects that as long as the price is within a factor two of the true price, no extra mean reversion term needs to be added to our dynamical equation for the price, Eq. (2.9). Mathematically, this mean reversion effect can be described by adding to the right hand side of Eq. (2.9) a term proportional to $-\kappa \log(P_t/P_0)$, where κ measures the strength of the demand driven by fundamental considerations and P_0 the true fair price. On short time scales, or if κ is sufficiently small, this term can be neglected and the analysis presented above should be valid. On long time scales, however, such that the random fluctuations become of the order of (say) 100 %, one should expect these mean reversion effects to become relevant (on this point, see also Föllmer et al. 2005). For a typical stock with a daily volatility of 3%, this corresponds to 1000 days, or four years. Such a time scale corresponds to the typical reversion time scale discovered by De Bondt and Thaler in their paper on overreaction in stock markets. Hence, in a world where absolute references are lacking, one expects that the short to medium time scale dynamics of markets will be dominated by the self-referential effects described in the present chapter.

3. Microstructure

We now focus on much shorter time scales, where the price formation takes place, and we study in particular continuous double-auctions. As we discussed in introduction, in this setting price formation is a self-organized process, where regularities and empirical power laws are observed. The goal is to understand some of these relations. In particular, what fixes liquidity, and how does it depends on volatility? After describing the continuous double-auction in section 3.1, we study some zero-intelligence models that were proposed to study it. In these models expressions of the spread and the volatility can be derived. Such models are helpful to get used to the "kinematics" of the problem, and underline important questions such as the very definition of price, that we also study empirically. Nevertheless we shall argue that they are far too naive to describe actual price formation. In section 3.3 we provide empirical results, in particular on the statistics of the market order flow, on the price diffusion and on the time behavior of the price impact function. One striking fact is that the market order flow (and, incidentally, the order-book itself) is a long-memory process. We show how to re-conciliate these long-memory correlations with the almost perfectly diffusive nature of price fluctuations. This implies that the impact of one single trade is almost completely *transient*. In section 3.4, we show that it is possible to express the gain of liquid orders versus market orders. This leads to a necessary relation between spread and price impact. We use this relation to show that volatility per trade and spread are proportional.

3.1 Continuous double-auction description

The description of the order-book below is taken from (Farmer, Gillemot et al. 2004). The double continuous auction is the standard mechanism for price formation in most modern financial markets. Agents can place different types of orders, which can be grouped into two categories: Impatient traders submit market orders, which are requests to buy or sell a given number of shares immediately at the best available price. More patient traders submit limit orders, or quotes which also state a limit price, corresponding to the

Fig. 3.1: A schematic illustration of the continuous double auction mechanism. Limit orders are stored in the limit order book. The arbitrary convention that buy orders are negative and sell orders are positive was adopted. As a market order arrives, it has transactions with limit orders of the opposite sign, in order of price (first) and time of arrival (second). The best quotes at prices $a(t)$ or $b(t)$ move whenever an incoming market order has sufficient size to fully deplete the stored volume at $a(t)$ or $b(t)$, when new limit orders are placed inside the spread (when the limit price p satisfies $b(t) < p < a(t)$) or when all the orders at the best prices are canceled. Bids and offers that fall inside the spread become the new best bids and offers. (Farmer, Gillemot et al. 2004).

worst allowable price for the transaction. (Note that the word quote can be used either to refer to the limit price or to the limit order itself.) Limit orders often fail to result in an immediate transaction, and are stored in a queue called the limit order book. Buy limit orders are called bids, and sell limit orders are called offers or asks. At any given time there is a best (lowest) offer to sell with price $a(t)$, and a best (highest) bid to buy with price $b(t)$. These are also called the inside quotes or the best prices. The price gap between them is called the spread $S(t) = a(t) - b(t)$. Prices are not continuous, but rather change in discrete quanta called ticks, of size δp. The number of shares in an order is called either its size or its volume. As market orders arrive they are matched against limit orders of the opposite sign in order of first price and then arrival time, as shown in Fig.(3.1). Because orders are placed for varying numbers of shares, matching is not necessarily one-to-one. For example, suppose the best order is for 200 shares at \$60 and the the next best is for 300 shares at \$60.25; a buy market order for 250 shares buys 200 shares at \$60 and 50 shares at \$60.25, moving the best order $a(t)$ from \$60 to \$60.25. A high density of limit orders per price results in high liquidity for market orders, i.e., it implies a small movement in the best price when a market order is placed. When a market order arrives it can cause changes in the best prices. This is called market impact or price impact. Note that the price changes are always in the same direction: a buy market order will either leave the best ask the same or make it bigger, and a sell market order will either leave the best bid the same or make it smaller. The result is that buy market orders can increase the midprice $m(t) = (a(t) + b(t))/2$, and sell orders can decrease it. Note that the arrival of three kinds of events can cause the midprice to change (i) Market orders. A market order bigger than the opposite best quote widens the spread by increasing the best ask if it is a buy order, or decreasing the best bid if it is a sell order (ii) Limit orders. A limit order that falls inside the spread narrows it by increasing the best bid if it is a buy order, or decreasing the best ask if it is a sell order. (iii) Cancellations. A cancellation of the last limit order at the best price widens the spread by either increasing the best ask or decreasing the best bid.

3.2 Zero-intelligence model

In the neoclassical approach of microstructure, price formation emerges from a strategic equilibrium between fully-rational agents, who are on the one-hand market makers who furnish liquidity, whose incomes come from the spread, on the other hand informed agents trying to use their information to forecast future price changes. From the empirical point of view, to our

knowledge the neoclassical approach has not yet being able to account for the linear relation between spread and volatility per trade (Zumbach, 2004; Kockelkoren, 2005), which stands as a fundamental question of the field.

Taking the opposite direction, some economists and physicists have proposed zero-intelligence models, where the behavior of agents is mimicked by random flows of limit and market orders. In this section we aim to present this approach. We discuss the general method and a few models, and justify the value of spread and volatility in this framework. This should provide intuition about order-book geometry to the reader unfamiliar with electronic markets. This also underlines interesting questions, such as the price definition. Nevertheless, we shall ultimately argue that this point of view is too naive to describe price formation. In particular zero-intelligence models miss the fact that imposing the absence of simple statistical arbitrage opportunities strongly constrain the order-book properties.

Zero intelligence models are in general not easily tractable analytically. These models were mostly studied numerically, and the numerical results were compared with empirical data. Here our goal is to furnish simple qualitative argument to rationalize the price dynamics they generate. Some of our results underline the inadequacy of these descriptions.

In the sections 3.3 and 3.4 our approach shall be more "data driven", and considers the constraints that the absence of statistical arbitrage impose to the price formation.

3.2.1 Modeling the order book

We shall denote the tick size—the elementary increment of price— δp (in dollars). In the simplest form of these models, the order flows are Poisson processes, and all orders are emitted with the same size γ. The market order flows (to buy and to sell) have identical rate μ (in shares per seconds). The flow rate of limit orders emitted on one tick is α (in shares per seconds per dollars). In the models we shall discuss, the regions where limit orders to buy and to sell fall is function of the instantaneous order-book geometry. We will see that the price dynamics depends much of this choice. Finally limit orders are cancelled randomly with a rate δ (per seconds).

In what follows we shall choose δp as the unit of price, δ^{-1} as the unit of time, and γ as the unit of volume of shares. Thus all results can be expressed in terms of the two unit-less parameters $\hat{\alpha} = \alpha[\delta p]\delta^{-1}\gamma^{-1}$ and $\hat{\mu} = \mu\delta^{-1}\gamma^{-1}$. For now on we shall drop the hats, and explore all the possible regimes of the following model by varying α and μ.

In what follows we use the notation "$c \sim d$" to indicate that $\lim_{d\to\infty} c/d \to C_0 \neq 0$, where C_0 is a numerical constant. In such circumstances we say that

c "scales" with d.

3.2.2 Maslov's model

Maslov (Maslov, 2000) proposed a simple model of order-book. This model is to simple too describe the order-book at the spread level, but displays curious price fluctuations at large time scales. It is defined as follows: at each time step, either a limit order is emitted with a probability 1/2, or a market order is emitted with the same probability. The order can be to sell or to buy, with a probability 1/2. When a market order to buy (to sell) is emitted, a transaction takes place at the bid (ask). If a limit orders is emitted, it falls at a price $P(t) + 1$ if it is a limit order to buy ($P(t) - 1$ if the limit order is to sell), where $P(t)$ is the last transaction price. If there are no more limit orders in the order-book, limit orders only can be emitted. Maslov studied this model without limit orders cancellation, that is $\delta = 0$. In our notation this corresponds to $\alpha = \mu = \infty$. Such a model displays surprising features: (i) the distribution $P(r)$ of the price return $r = \frac{p(t+1)-p(t)}{p(t)}$ has fat tails: $P(r) \sim r^{-(1+x)}$. ii) there are long range correlation in volatility and (iii) the price sub-diffuses as $t^{1/4}$. Maslov noted that (i) and (ii) are indeed observed in the price fluctuations, while (iii) is wrong: the price in fact diffuses as $t^{1/2}$. Nevertheless, this behavior is related to unrealistic market assumptions. In Maslov's model, price formation is not a stationary process. The order-book displays in fact aging: in particular, it is straightforward to show that the volume contained in the order-book follows a random walk. Therefore this volume grows as $t^{1/2}$, and the order-book never reaches any stationary limit.

3.2.3 Farmer's model

Farmer et al. (Farmer, Patelli, et al. 2004) proposed a model belonging to the same class, but with more realistic rules. This enables to generate more reasonable order-books, spreads and to recover a diffusive behavior at *large* time scales. The model is identical to Maslov's, except that: (i) cancellations do occur and (ii) the deposition rule of limit order is different. Limit orders to buy fall uniformly on all ticks of price p such that $p < a(t)$. Reciprocally, limit order to sell fall uniformly for $p > b(t)$. Note that (i) limit orders of both types fall inside the spread. In fact the limit order deposition is symmetrical around the mid-point $m(t) = \frac{1}{2}(a(t) + b(t))$, whereas in Maslov's model this symmetry is around the last transaction price. (ii) the limit order flow does not decay far away from the price, as it does in real market (Bouchaud, Potters & Mèzard, 2002). Nevertheless this issue is not relevant, since in this model limit orders far away from the price are in general cancelled before

they can participate to a transaction. Thus they do not affect the price dynamics.

The dynamics generated by this model is in fact quite complex and non-linear. In particular, each time a limit order falls inside the spread, the region of deposition for the limit orders of the other sign is changed. We shall see that it has very strong effect. A mean-field theory was proposed (Smith et al., 2003), which studies the order-book in the mid-point referential. This approximation does not work well in general, notably to explain the price diffusion D, or volatility square, which scales (i.e varies as a power-law for large μ) in the simulation as $\mu^{3/2}$ for a large domain of parameters. In what follows we show that this approach fails because the mid-point is not the good referential to consider. Furthermore, we argue that the peculiar dependence of volatility with the market order flow is not a robust property of the model, and is generated by an unrealistic choice of limit order deposition.

Spread and order-book depth

We define the order-book *depth* $h(p)$ as the volume of limit orders at price p. For all parameters regimes, the depth $h(\infty)$ of the order-book far away from the price results from the balance between the flow of limit orders and cancellations, since market orders are never executed there. This balance yields on average:

$$\frac{dh}{dt} = \alpha - h \tag{3.1}$$

Thus at equilibrium $h = \alpha$. Note that Eq.(3.1) also shows that the "relaxation time" of the order-book is $\delta^{-1} \equiv 1$: if there is a small fluctuation from the average depth value $h(0) = \alpha + \delta h_0$, then $h(t) - \alpha = \delta h_0 \exp(-t)$.

Another simple result concerns the dependence of the average spread $S = \langle S(t) \rangle$. Inside the spread, the limit order flow $S\alpha$ is compensated by cancellation and market orders. When market orders dominate cancellations, that is when $\mu \gg 1$, one finds (Farmer, Patelli, et al. 2004) $S\alpha = \mu$, or:

$$S \sim \frac{\mu}{\alpha} \tag{3.2}$$

See Fig.3.4 for a numerical simulation of the corresponding order-book. When cancellation dominates $\mu \ll 1$ and the limit order rate is small $\alpha \ll 1$, one finds that the spread goes like the distance between two limit orders of same sign $1/h$, see Fig.3.3 for a numerical simulation:

$$S \sim \frac{1}{\alpha} \tag{3.3}$$

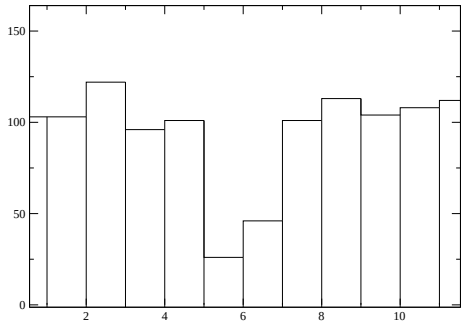

Fig. 3.2: Snapshot of the depth profile of a simulated order-book *vs.* price for $\alpha = 100$, $\mu = 50$. The mid-point is 6. The bids are below 6, the offers are above.

Note that when $\alpha \gg \mu$ and $\alpha \gg 1$, the average spread S is minimum and equal to 1 tick: $S = 1$, see Fig.3.2.

These results assume that the deposition of limit orders is constant inside the spread. Nevertheless empirically this is not at all the case (Farmer, Gillemot et al., 2004). In particular, there are roughly as many limit orders to sell falling inside the spread than exactly at the ask, in disagreement with the model assumption as soon as the spread is larger than one tick. Thus there is no reason to think that spread and order-book depth are related by Eq.(3.2), as proposed by (Farmer, Patelli & Zovko, 2004). Empirical results (Zumbach, 2004) support that it is indeed not the case [1].

In what follows we study the more subtle question of the volatility dependence for different regions of parameters.

$\alpha \gg \mu$, $\alpha \gg 1$: Limit order dominated

In this parameters domain $S = 1$, and the order-book shape is as follows: the volumes at $b(t)$ and $a(t)$ are on average $\alpha - \mu$, elsewhere the order-book depth is on average α, see Fig.(3.2). The price changes are rare, they take place when the volume at the bid or at the ask fluctuates down to zero. When this happens, with a probability 1/2 the price changes of one tick, otherwise the price recovers its previous value. Thus in the domain of parameters the diffusion constant defined as:

$$D \equiv \langle \frac{1}{t}(p(t) - p(0))^2 \rangle \equiv \sigma^2 \tag{3.4}$$

[1] The characteristic depth of the order-book in unit of value per transaction is, according to Zumbach, $h = \alpha \sim C^{0.76} C^{-0.44} \sim C^{0.3}$. The frequency of trade follows $\mu \sim C^{0.39}$. Applying Eq.(3.2) leads the to $S \sim C^{0.09}$ instead of the observed $S \sim C^{-0.25}$.

follows $D \sim \tau^{-1}$, where τ is the average time between two changes of price. A similar result applies to a discrete random walk which makes jumps of amplitude 1 every time τ. In what follows we aim to evaluate the order of magnitude τ. To do so we consider the total volume of limit order at the bid (or at the ask). We denote by f_k the probability of having a volume k. The probability per time unit to change price is $\frac{1}{2}f_0\mu$, thus $\tau = 2\mu^{-1}f_0^{-1}$. To determine f_k, we write a master equation. A master equation lists the probability for a volume k to increase by 1 when a limit order is emitted (this happens at a rate αf_k), and the probability to decrease by one when a market order is emitted, or when a cancellation occurs (rate $f_k(\mu + k)$). By taking into account all these events, it is possible to write an equation for the evolution of probabilities $f_k(t)$ for all k:

$$\frac{df_k}{dt} = (\mu + (k+1))f_{k+1} - (\mu + \alpha + k)f_k + \alpha f_{k+1} \text{ for } k \leq 1 \quad (3.5)$$

$$\frac{df_0}{dt} = (\mu + 1)f_1 - \alpha f_0 \quad (3.6)$$

It is possible to solve such problem by introducing the generating function $H(x) = \sum_{k=1}^{\infty} x^k f_k$ we obtain:

$$(1-x)H'(x) + (\frac{\mu}{x} - (\mu + \alpha) + x\alpha)H(x) - (1+\mu)(x-1)f_1 = 0 \quad (3.7)$$

$$H(1) = 1 - f_0 = 1 - \frac{\mu + 1}{\alpha}f_1 \quad (3.8)$$

which can be solved to obtain:

$$H(x) = f_1(1+\mu)\int_0^x \left(\frac{t}{x}\right)^{\mu} e^{\alpha(x-t)}dt \quad (3.9)$$

and

$$f_1 = [\frac{\mu+1}{\alpha} + (1+\mu)\int_0^1 t^{\mu}e^{\alpha(1-t)}dt]^{-1} \quad (3.10)$$

We use the saddle point method to approximate the integral. This approximation is used when large exponential pre-factors are involved, as it is the case here since $\alpha \gg 1$. We find, neglecting the pre-factors in comparison with the exponential term:

$$D \sim (\mu f_0) \sim f_1 \asymp \exp[\alpha - \mu + \mu \ln(\frac{\mu}{\alpha})] \quad (3.11)$$

where the notation $c \asymp d$ means $\frac{\ln(c)}{\ln(d)} \to 1$ as $d \to \infty$. In practice, there are stocks for which $S = 1$, they are said to have "big" ticks. This is the case for example for the Swedish market. In this situation, as we showed here, this

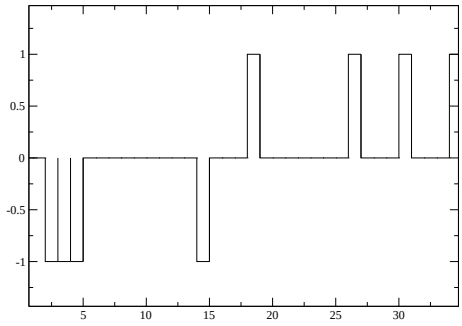

Fig. 3.3: Snapshot of the order-book depth profile *vs.* price for $\alpha = 0.2$, $\mu = 0.1$. Asks are presented with positive values, bids are presented with negative values.

model leads to a very low volatility: an "activated event", that is a rare fluctuation, must takes place for the price to move. This leads to exponentially large time scale and small volatility. Neverthless, this phenomenon is not observed in real markets, where the price does not "stick" to discrete values. In particular, empirically the probability $P(V)$ that the volume at the bid (or at the ask) at small V is not small. This suggests that the agents do not consider only the midpoint as an indicator to place their limit orders, but also the volume at the bid and at the ask. It is possible to propose models which use this idea to recover a normal, non-activated dynamics.

$\mu \ll 1$, $\alpha \ll 1$: Cancellation dominated markets

In this case, even far from the price the order-book depth is smaller than 1 on average, see Fig.(3.3). Because $\mu \ll 1$, the diffusion is dominated by the cancellation and the emission of limit orders. Cancellation occurs with a characteristic time equal to unity, and the jumps of price are of the order of the spread α^{-1}. It follows that:

$$D \sim \frac{1}{\alpha^2} \tag{3.12}$$

$1 \ll \mu$, $\alpha \ll \mu$: Market order dominated markets

This is the most subtle case. In what follows we shall only sketch an explanation for the volatility, and explain why the referential of the mid-point is not natural. The diffusion constant D found numerically (Farmer, Patelli,

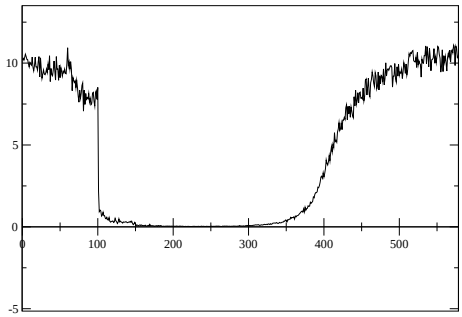

Fig. 3.4: Average order-book depth *vs.* price in the referential of b^* for $\alpha = 10$, $\mu = 2000$. Note that in this referential the order-book depth is not symmetric: the interface b^* is much sharper than in a^*. The interface in b^* is similar to the "typical interface" that one can observe by taking a snapshot of the depth profile, without averaging. By contrast, the interface in a^*, seen from b^*, is smoothed by the fluctuations of the effective spread S^*, which represents the distance between the two interface.

2004) is:

$$D \sim \frac{\mu^{3/2}}{\alpha^2} \tag{3.13}$$

This is a surprising result. A naive argument gives a different diffusion constant. Indeed, the order-book receives market orders, or "shocks", with a frequency μ. A volume V of market order moves one side of the order-book by an amount of roughly $\frac{V}{h}$, where h is the characteristic order-book depth. We may take for h the value of the depth of the order-book at infinity $h = \alpha$. Thus the mid-point changes by an amount $\frac{V}{2\alpha}$. In this simple picture, each unit market order changes the price by an amount $\frac{1}{2\alpha}$. This price dynamics is then equivalent to a Brownian motion where shocks appear with a rate μ and an amplitude $\frac{1}{2\alpha}$, which leads to $D \sim \frac{\mu}{\alpha^2}$, in contradiction with the simulations.

A direct numerical inspection of the order-book shape shows that, in this regime of parameters, a "slow" and persistent structure appears in the order-book. The order-book displays two interfaces, see Fig.(3.4). These interfaces delimits a region where the order-book depth is nearly the depth at infinity $h = \alpha \gg 1$ from a region where the depth is suddenly much lower. We shall denote $a^*(t)$ and $b^*(t)$ the position of these two interfaces, defined for example as the prices where the order-book depth become larger than half the depth at infinity [2]. It is convenient to introduce an effective spread

[2] For simplicity we shall consider $\alpha \geq 1$, but the same argument can be easily extended

$S^* = a^* - b^*$ and an effective mid-point $m^*(t) = 1/2(a^*(t) + b^*(t))$. Fig 3.4 shows the average order-book in the referential of $b^*(t)$. This structure moves much more slowly than the mid-point $m(t)$ itself: the mid-point makes jumps of the order the spread with a time scale μ^{-1}. This observation indicates why trying to describe the order-book in the referential of $m(t)$ fails: it is inadequate to describe a slow-moving structure in a fast-moving referential.

We aim to describe the motion of this slow structure. First we evaluate S^*. We introduce $r_a = a^* - a$ and $r_b = b - b^*$, the distances between effective ask and ask, and between effective bid and bid. By symmetry r_a and r_b have equal distributions. We have $S^*(t) = S(t) + r_a(t) + r_b(t)$. Now we repeat an argument of flow balance: the flow of limit orders to sell falling between b and a^* must compensate the market orders to buy. Although b displays fast fluctuations, this balance must be verified on average. Thus we obtain $\langle r_a \rangle + S = \frac{\mu}{\alpha}$. Therefore $S^* = \frac{\mu}{\alpha} + \langle r_a \rangle$. To evaluate $\langle r_a \rangle$ we shall provide a simple model with one interface. It predicts $\langle r_a \rangle \approx \frac{\mu}{2\alpha}$. Thus we find, in good agreement with the simulations, see Fig.3.4:

$$\langle S^*(t) \rangle \approx \frac{3}{2} \frac{\mu}{\alpha} \tag{3.14}$$

Thus the spread S is a fraction (roughly 2/3) of S^*. Since the limit orders density is small in S^*, S rapidly diffuses on the segment S^*. The motion of $m(t)$, the center of S, is therefore strongly anti-correlated: S "bounces" back on the two interfaces a^* and b^*. Such strong (and as we shall comment later, unrealistic) anti-correlations are observed numerically (Farmer, Patelli et al., 2004).

This suggests the following explanation for the surprisingly large long-time diffusion, corresponding to the motion of the slow structure: the fluctuations of S create fluctuations of the limit order flow, much larger than the fluctuation in market orders. We shall denote τ_S the characteristic time of the fluctuations of S in S^*, defined for example as the average time for S to go from one interface to the opposite one. τ_S is to be determined . Fig.(3.5) shows a snapshot of the order-book where the spread is bouncing on the interface a^*. During the time τ_S, the imbalance between the two opposite limit order flows falling in S^* is of the order of $\tau_S S\alpha = \tau_S \mu$. This imbalance leads directly to an imbalance on the market orders which impact the interfaces. Since on average each market order which reach an interface changes the interface position by an amount α^{-1}, during τ_S the interfaces center moves by a random amount of the order $\tau_S \mu/\alpha$. This leads to a diffusion constant

to the region $\alpha \leq 1$, with more care for the definition of a^* and b^*

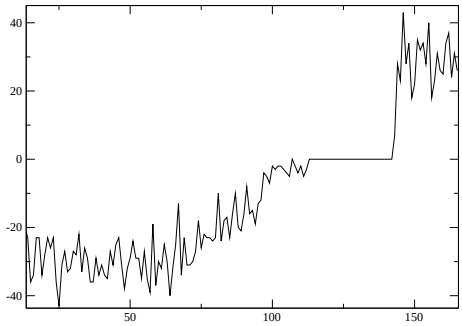

Fig. 3.5: Snapshot of the order-book depth *vs.* price for $\alpha = 30$, $\mu = 1000$. Asks are presented with positive values , bids with negative values. We have $b^* = 95$ and $a^* = 143$. For $p \in [b^*, a^*]$, there are around 30 bids present, and no asks.

scaling as:

$$D \approx \left(\frac{\mu\tau_S}{\alpha}\right)^2 \frac{1}{\tau_S} = \frac{\mu^2\tau_S}{\alpha^2} \tag{3.15}$$

Following (Farmer, Patelli et al., 2004) we may also define a diffusion constant at small time scales D_0, which describes the sub-diffusive motion of the mid-point at short times $t \ll \tau_S$. Since in a time τ_S a distance of order S is covered, one obtains:

$$D_0 \sim \frac{\mu^2}{\alpha^2\tau_S} \sim D/\tau_S^2 \tag{3.16}$$

The numerical results of (Farmer, Patelli et al., 2004) for D and D_0 are consistent with $\tau_S = \mu^{-1/2}$. In this paragraph we propose a hand-waving argument for the dependence of τ_S. Consider that there are typically N limit orders inside S^*. The time scale to remove one order by a market order is μ^{-1}, so that $\tau_S \sim N\mu^{-1}$. The fluctuations of the market order flow during the relaxation time 1 of the order-book time are of order $N \sim \mu^{1/2}$. This leads to $\tau_S = \mu^{-1/2}$ as observed numerically. This rationalizes the scaling of both D and D_0.

Relation with real markets

Here we argue that the present model is inadequate to describe market volatility. In particular, the peculiar scaling of the diffusion constant $D \sim \frac{\mu^{3/2}}{\alpha^2}$ comes from a subtle non-linear effect: when limit orders fall inside the spread, they affect the deposition of the limit orders of the other signs. This effect is not robust against changes in the rules of spatial repartition of limit orders. In real markets, as we said above, there are roughly the same number

of limit orders deposited on the bid and ask than inside the spread. In the present model, if the deposition flow of limit orders inside the spread decreases, the non-linear effect disappears and D becomes proportional to μ. Furthermore, the signature of this non-linear effect is the anti-correlation of the mid point. This is not observed on real market, where the mid-point dynamics is well-described by pure diffusion, even at short time (for small tick size).

The present model underlines the interesting question of the very price definition, that we shall study empirically in what follows. Here the zero-intelligent agents emit limit orders symmetrically around the mid-point m. Doing so they do not realize that the dynamics of m is highly anti-correlated, and that m fluctuates rapidly around m^*, which represents the slow "center of mass" of the order-book. In this model $m^*(t)$ is a much better estimate of the future mid-point than the instantaneous mid-point itself. Rational agents should use this information to place their orders. In other words, in the present model simple statistical arbitrage opportunities exist. For example, since some limit orders to buy are deposited above m^*, and some limit orders to sell are deposited above, a simple strategy with positive profit is to emit market orders against all these limit orders.

In the next paragraph we present, for completeness, an alternative model where limit orders are placed symmetrically around m^*. In this model the arbitrage discussed above is not possible, and a normal diffusion constant $D \sim \frac{\mu}{\alpha^2}$ is recovered.

3.2.4 Alternative model with effective price

The model we propose is identical to the previous one, with the only difference that limit orders are now deposited relatively to the effective mid-point m^*. Limit order to buy flows with a rate α for $p \geq m^*$, the symmetric rule defines the flow of limit orders to sell. We focus on the case where $\alpha \ll \mu^{1/2}$, $\mu \gg 1$. As we shall see this model is simpler than the previous one. In the referential of m^*, it corresponds to the interaction of two distinct interfaces whose properties are only determined by the arrival of limit orders of one single sign, and market orders of the other sign.

Problem with one interface

We consider the dynamics of the upper part of the order-book (the offer) in the referential of m^*. It is, at a good level of approximation, equivalent to fix $m^* = 0$, and to consider the price segment $[0, \infty]$. Limit order flow at each price unit with a rate α. Market order flow with a rate μ. When a

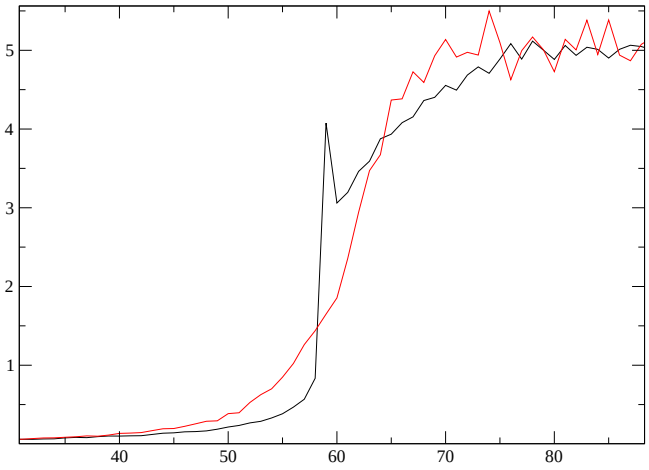

Fig. 3.6: The black curve is the average depth profile *vs* price in the referential a^*. The red curve is the average depth profile in the non moving referential. These curves correspond to $\alpha = 5$ and $\mu = 300$.

market order is emitted, the transaction takes place for the limit order with the lowest price. Limit orders are cancelled with a rate 1. A typical interface is shown in Fig.(3.6).

A simple balance of flows gives for the average position of the interface $\langle a^* \rangle \approx \frac{\mu}{\alpha}$. It is also possible to furnish a qualitative argument to evaluate the fluctuations of a^*. As we discussed in Eq.(3.1), the order-book relaxes with a characteristic time equal to 1. Thus the fluctuations of a^* are given by the fluctuations of the number of market orders during that time interval, divided by the order-book depth above m^*, which is approximately α. This lead to $\langle (a^* - \langle a^* \rangle)^2 \rangle^{1/2} \sim \frac{\sqrt{\mu}}{\alpha}$.

Furthermore, we find that the average position r_a of the first limit order is $r_a = C_0 \frac{\mu}{\alpha}$, where $C_0 \approx 1/2$. In Appendix 1 we justify further these results using simple approximations.

Price diffusion

During a time unit, each interface has moved randomly of an amount $\Delta \sim \frac{\sqrt{\mu}}{\alpha}$. Thus Δ characterizes both the fluctuation of the effective spread, but also the fluctuation of the effective mid-point position. In a good approximation price follow a random walk with random steps of size Δ per time unit. This leads to a diffusion constant:

$$D \sim \Delta^2 \sim \frac{\mu}{\alpha^2} \tag{3.17}$$

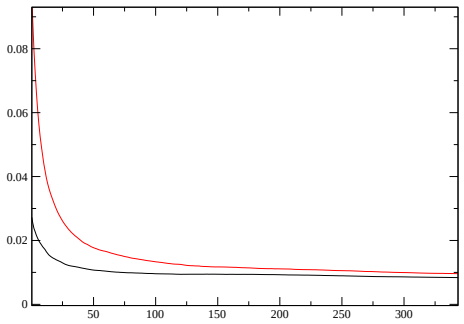

Fig. 3.7: $D_x(\ell) \equiv \frac{1}{\ell}\langle (x(\ell) - x(0))^2 \rangle$ *vs.* ℓ for $x = m$ (red curve) and $x = m^*$ (black curve). Simulations were done with $\mu = 12$ and $\alpha = 6$.

It is interesting to consider in more details the diffusion of the effective mid-point m^* and of the mid-point m in this model. We compute the quantity $D_x(\ell) \equiv \frac{1}{\ell}\langle (x(\ell) - x(0))^2 \rangle$, where ℓ is the number of transaction that took place during the time interval t. When $\ell \to \infty$, D_x converges to the diffusion constant. We compute this quantity for $x = m$ and $x = m^*$. Fig.(3.7) shows that both m^* and m are sub-diffusive. Nevertheless this effect is much smaller for m^*. By contrast, m sub-diffuses more: this is the signature of the fast anti-correlated motion of m, which oscillates around the center of mass of the order-book m^*. Note that the simple statistical arbitrage that was possible in the model of Farmer et al. is very reduced here, because the limit orders to buy (to sell) are emitted above (below) m^*, which is a good estimator of the future mid-point.

3.2.5 Discussion

To conclude on this subject, we argue that zero-intelligence models are not a good starting point to study price formation. The reason is that, as we shall see in section 3.4, forbidding the existence of simple statistical arbitrages much constrains the order-book and the price dynamics. Zero-intelligence models completely miss this point. In particular, such models do not lead to a correct estimation of the order-book properties, such as the spread. These models were also used to justify the concavity of the price impact. The argument uses that in these models the average depth profile increases away from the price, before saturating toward a constant. Thus, when a market order with a large volume V is emitted, its impact on price $\mathcal{R}(V)$ is sub-linear with V. More precisely, assuming that the order-book equal its

average value, the impact function satisfies:

$$V = \int_0^{2\mathcal{R}(V)} h(p)dp \tag{3.18}$$

For example if the order-book average depth increases linearly, this argument would yield a square root for the price impact. Nevertheless it turns out empirically that the cause of the impact function concavity is different. In practice there are almost no market orders that "penetrate" in the order-book (Farmer et al., 2004), that would lead to transaction further from the bid or the ask. The concavity of the price impact has rather to do with a strong positive correlation between market orders volume and the limit orders volume at the bid, or at the ask.

Such models are not realistic either to compute the price dynamics and the volatility. As we shall see, the price diffusion emerges from a subtle compensation of two strong antagonist effects: the long-memory of the market order flow, and the long memory of the order-book. Such effects are not contained in zero-intelligence models.

These models could certainly be improved by adding more and more realism, for example by conditioning the order flow to the past, or to details of the order-book shape. Nevertheless this would not do better than reproducing real markets. Zero-intelligence approach could bring insights to the price formation only if the regularities observed on financial market stemmed from zero-intelligence, or complete disorder. Simple arbitrage strategies are sufficient to invalidate this hypothesis.

3.2.6 Empirical results on "effective" price

One interesting outcome of zero-intelligence models is that the mid-point $m(t)$ is not the most appropriate definition of price. It does not take into account the whole order-book shape, but simply the bid and ask positions. We may define rigorously the "efficient" price $m_{eff}(t)$ as the best estimate of the future mid-point knowing the (observable) instantaneous order-book: $m_{eff}(t)$ is defined such that $\langle (m(t') - m_{eff}(t))^2 \rangle$ is minimum when $t' - t$ is large (say of the order of the hour). In our zero-intelligence model, m_{eff} is roughly at the middle of the two interfaces $m_{eff} \approx m^* \equiv \frac{1}{2}(a^* + b^*)$. It is reasonable to think that on real markets also, one may find a better estimate of the future mid-point than the mid-point itself.

In what follows we test this hypothesis and propose a simple estimation of m_{eff} that uses the order-book shape. We show that there exists indeed a better estimation of the future mid-point that the mid-point itself. It turns out to be a linear combination of the mid-point m and the center of mass of

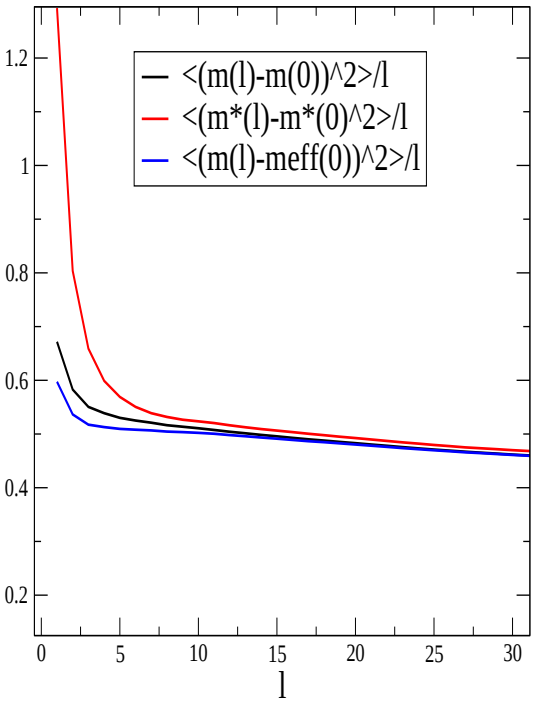

Fig. 3.8: Plots of $\frac{1}{\ell}\langle[m(\ell)-m(0)]^2\rangle$, $\frac{1}{\ell}\langle[m(\ell)-m_{eff}(0)]^2\rangle$ and $\frac{1}{\ell}\langle[m^*(\ell)-m^*(0)]^2\rangle$ as defined in the text *vs* the number trades ℓ for Astrazeneca.

the order-book m_c^*, that we shall define more precisely for real order-books in what follows. Nevertheless, we show empirically that there are importance differences between the center of mass in real order-book m_c^* and in zero intelligence models: in particular we find that m_c^* fluctuates *more* than the mid-point m. We also show that $m - m_c^*$ furnish indications on the future market order flow.

We consider the five best bid prices and ask prices where limit orders are present (only these can be observed in electronic markets). We introduce a^* and b^*, the center of mass of limit orders to buy and to sell. For example, $a^* = \frac{1}{V_a}\sum_{i=1}^{5} V_i p_i$, where V_a is the total volume of limit orders to sell observable, V_i the volume of limit orders at the i^{th} best ask, whose price is p_i. Then, we define m_c^* as $m_c^* = \frac{1}{2}(a^* + b^*)$.

Then we consider the estimator $m_{eff} = ym + (1-y)m_c^*$. For Astrazeneca, by choosing $y = 0.8$, we show that m_{eff} is a better estimator of the future mid-point than the mid-point itself. Indeed we find that $\langle[m(\ell)-m(0)]^2\rangle < \langle[m(\ell)-m_{eff}(0)]^2\rangle$ as can be seen in Fig.3.8. ℓ is the lag in trade time.

In our zero-intelligence model, the dynamics of m^* correspond to the

slow motion of the center of mass of the order-book, whereas m has fast, anti-correlated motions inside the spread. Thus, m^* estimates better the future mid-point than m, which "oscillates" around m^*. As a consequence, m presents an anti-persistence, or sub-diffusive, effect much larger than the one of m^*, which is almost diffusive, as it is apparent in the numerical result of Fig.(3.7). To test if this scenario is correct to describe real order-books, we compute $\frac{\mathcal{D}(\ell)}{\ell} = l^{-1}\langle(x(\ell) - x(0))^2\rangle$ for $x = m_c^*$ and $x = m$. The results are shown in Fig.(3.8). In stark contrast with the zero-intelligence model, we observe that m_c^* is more sub-diffusive than m.

We found empirically another interesting difference between real markets and zero-intelligence models: $m_{eff} - m$ is a good predictor because it furnishes an information on the future flow of market orders, in contrast with the previous models where the market order flow is purely random. Fig.(3.9) shows the correlation $C(\ell) = \langle\epsilon(\ell)(m(0) - m_{eff}(0))\rangle$, where $\epsilon(\ell)$ is the market order sign, chosen positive if the market order is to buy. $C(t)$ is indeed found to be negative at short time (up to 7 transactions time). A possible explanation for this is that when a piece of new arrives on the market, some liquidity providers remove their limit orders before market orders can be emitted. This would lead to the observed correlation.

To conclude, our empirical results show that, as it is the case for zero intelligence models, the very definition of price is rather subtle. It is possible to define an efficient price which predicts better the future mid-point than the mid-point itself. This efficient price takes into account the order-book shape. Nevertheless the analogy with zero-intelligence models is not perfect. In such models the efficient price is simply the mass center of the order book, around which the mid-point oscillates. In contrast, the center of mass of the order-book fluctuates more than the mid-point in real market. Furthermore the effective price contains another information, on the flow of incoming market orders.

3.2.7 Appendix: Computation of the average depth profile of one single interface

It is possible to compute exactly the average order-book depth of one single interface, shown in Fig.3.6, as follows: consider the first p ticks. In this interval, the rate of the total limit order flow is αp. In our previous argument which follows Eq.(3.5), we computed the volume distribution for one tick with a limit order rate α. We can use the same argument here by changing α in αp to compute the average number of limit orders $k(p)$ located in the interval

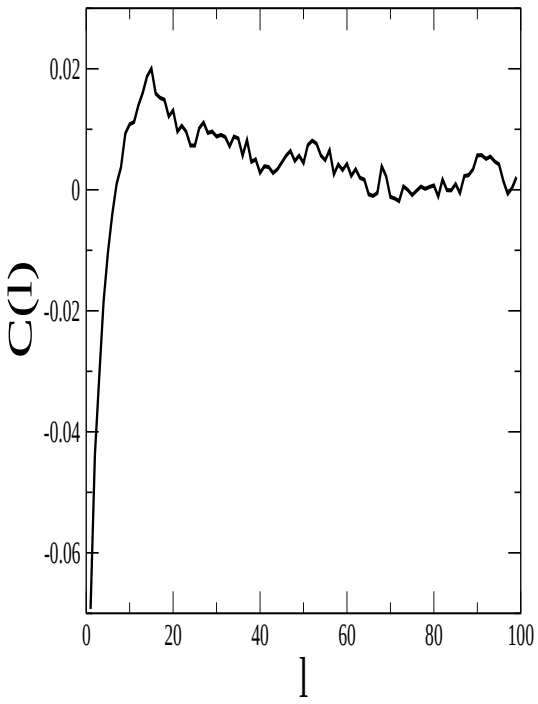

Fig. 3.9: Plot of $C(\ell) = \langle \epsilon(\ell) \cdot (m(0) - m_{eff}(0)) \rangle$ *vs* ℓ for Astrazeneca. The correlation is negative for $l < 7$.

$[0, p]$. Then the average depth $h(p)$ of the order-book would simply follow:

$$h(p) = \frac{dk}{dp} \tag{3.19}$$

In the framework that follows Eq.(3.5) we have:

$$k(p) = \sum_{k=1}^{\infty} f_k k = \frac{dH}{dx}|_{x=1} \tag{3.20}$$

where α must be replace by αp in the expression of H. Then Eqs.(3.9-3.10) enables to compute exactly $k(p)$. These equations can be solved numerically to obtain the quantities we are looking for. Here we rather provide simple approximations.

We estimate first the fluctuations of a^*. Fig.(3.6) shows the simulated average and typical depth profile, which was computed by averaging the order-book in the referential of a^*. The average profile is much smoother than the typical interface, because it is softened by the fluctuations of a^*. To evaluate these fluctuations in the regime $p > \mu/\alpha$, we note that the probability for $a^* > p$ is related to $f_0(p)$: if the interface is further than p, market orders must have reached at least p, which is possible only if there was no limit order at lower price. Thus we expect the characteristic decay length of $f_0(p)$ as p increases to describe the fluctuation of the interface position. From Eq.(3.11), after few lines of calculations, on obtains that at the leading order:

$$f_0 \asymp (\frac{1}{\mu}^{1/2}) \exp[-\frac{(p-\mu/\alpha)^2\alpha^2}{2\mu}] \tag{3.21}$$

which shows that the characteristic length $L = p - \alpha/\mu$ above which $f_0(p)$ vanishes follows $L \sim \frac{\sqrt{\mu}}{\alpha}$. From this we can deduce that $\langle (a^* - \langle a^* \rangle)^2 \rangle^{1/2} \sim \frac{\sqrt{\mu}}{\alpha}$ as stated.

We now evaluate r_a. At small $p \ll \frac{\mu}{\alpha}$, the probability $f_n(p)$ to have n limit orders in the interval $[0, p]$ satisfies $f_n(p) \ll f_{n-1}(p)$. Thus we may neglect f_n for $n \geq 2$. Using Eq.(3.5), and $f_0(p) + f_1(p) = 1$, we obtain $f_0(p) = \frac{1}{1+p\alpha/\mu}$ and $f_1 \approx \frac{\alpha p}{\mu}$. In this approximation the probability that $r_a > p$ is $P(r_a > p) = f_0(p)$, thus $P(r_a = p) = \alpha/\mu$, which leads to $r_a = 1/2$. This results approximates well numerical results.

3.3 The long-memory of financial markets

In this section we provide empirical data to show that the market order flow is a long-memory process. The market order signs ϵ are correlated in

time. This correlation decays very slowly, as a non-integrable power law: $\langle \epsilon(\ell)\epsilon(0)\rangle \sim \ell^{-\gamma}$ with $\gamma < 1$. It is observable for days. It seems paradoxical that such temporal correlations do not lead to a super-diffusive price dynamics. Indeed if the market orders impact were permanent, one would obtain $\frac{1}{\ell}\langle (P(\ell)-P(0))^2\rangle \sim \ell^{1-\gamma}$. We want to argue, based on a series of detailed empirical results obtained on trade by trade data, that the random walk nature of prices is in fact highly non trivial and results from a fine-tuned competition between two populations of traders, liquidity providers (who emit limit orders) on the one hand, and liquidity takers (who emit market orders). For reasons that we explain in more details below, liquidity providers act such as to create anti-persistence (or mean reversion) in price changes that would lead to a sub-diffusive behavior of the price, whereas liquidity takers' action leads to long range persistence and super-diffusive behavior. Both effects very precisely compensate and lead to an overall diffusive behavior, at least to a first approximation, such that (statistical) arbitrage opportunities are absent, as expected. However, one can spot out the vestiges of this subtle compensation from the temporal structure of the market impact function. A surprising outcome of our work is that the response to one single trade decays slowly in time, and is almost *completely transient*.

The organization of this section is as follows. We first present (subsection 3.3.1) our empirical results on the statistics of trades, market impact and fluctuations. We show in particular that the order flow exhibits long range (power-law) autocorrelations in time, but that this does not lead to any predictability in price changes, as also recently noticed (Hopman, 2002). Then, we introduce in Subsection 3.3.2 a simple model that expresses the price as a linear superposition of the impact of each trade. We show that this model allows to rationalize our empirical findings, provided a specific relation between the temporal autocorrelation of the sign of the trades (i.e. buyer initiated or seller initiated) and the temporal response to a single trade is satisfied. Finally, in subsection 3.3.3, we give intuitive arguments that allow one to understand the market forces at the origin of this subtle balance between two opposite effects, which dynamically leads to absence of statistical arbitrage opportunities.

3.3.1 Market impact and fluctuations

Presentation of the data and definitions

In this study, we have analyzed trades and quotes data from liquid French stocks in the years 2001 and 2002, although qualitatively similar results were also obtained on British stocks as well. The advantage of the French market,

however, is that it is fully electronic whereas only part of the volume is traded electronically in the London stock exchange. We will illustrate our results mainly using the France-Telecom stock, which is one of the most actively traded stocks, for which statistics are particularly good.

There are two data files for each stock: one gives the list of all successive *quotes*, i.e. the best buy (bid, b) and sell (ask, a) prices, together with the available volume, and the time stamp accurate to the second. A quote can change either as a result of a trade, or because new limit orders appear, or else because some limit orders are cancelled. The other data file is the list of all successive *trades*, with the traded price, traded volume and time stamp, again accurate to the second. Rarely, several trades are recorded at the very same instant but at different prices: this corresponds to a market order of a size which exceeds the available volume at the bid (or at the ask), and hits limit orders deeper in the order book. In the following, we have grouped all these trades together as a single trade. This allows one to create chronological sequences of trades and quotes, such that between any two trades there is at least one quote.

The last quote before a given trade allows one to define the sign of each trade: if the traded price is above the last midpoint $m = (a + b)/2$, this means that the trade was triggered by a market order (or marketable limit order) to buy, and we will assign to that trade a variable $\varepsilon = +1$. If, one the other hand the traded price is below the last midpoint $m = (a + b)/2$, then $\varepsilon = -1$. With each trade is also associated a volume V, corresponding to the total number of shares exchanged.

Trades appear at random times, the statistics of which being itself non trivial (there are intra-day seasonalities and also clustering of the trades in time). We will not be interested in this aspect of the problem and always reason in terms of trade time, i.e. time advances by one unit every time a new trade (or a series of simultaneous trades) is recorded. We have also systematically discarded the first ten and the last ten minutes of trading in a given day, to remove any artifacts due to the opening and closing of the market. Many quantities of interest in the following are two-time observables, that is, compare two observables at (trade) time n and $n + \ell$. In order to avoid overnight effects, we have restricted our analysis mostly to intra-day data, i.e. both n and $n + \ell$ belong to the same trading day. We have also assumed that our observables only depend on the time lag ℓ.

On the example of France-Telecom, on which we will focus mostly, there are on the order of 10 000 trades per day. For example, the total number of trades on France-Telecom during 2002 was close to $2.\,10^6$; this allows quite accurate statistical estimates of various quantities. The volume of each trade was found to be roughly log-normally distributed, with $\langle \ln V \rangle \simeq 5.5$ and a

root mean square of $\Delta \ln V \simeq 1.8$. The range of observed values of $\ln V$ is between 1 and 11.

Price fluctuation and diffusion

The simplest quantity to study is the average mean square fluctuation of the price between (trade) time n and $n + \ell$. Here, the price p_n is defined as the mid-point just before the nth trade: $p_n \equiv m_{n^-}$. In this section, we always consider detrended prices, such that the empirical drift is zero. We thus define $\mathcal{D}(\ell)$ as:

$$\mathcal{D}(\ell) = \left\langle (p_{n+\ell} - p_n)^2 \right\rangle . \tag{3.22}$$

$\mathcal{D}(\ell)$ is the diffusion per transaction time, and must be distinguished from the diffusion per unit time D discussed in the previous section. As is well known, in the absence of any linear correlations between successive price changes, $\mathcal{D}(\ell)$ has a strictly diffusive behavior, i.e.

$$\mathcal{D}(\ell) = D\ell, \tag{3.23}$$

where D is a constant. In the presence of short-ranged correlations, one expects deviations from this behavior at short times. However, on liquid stocks with relatively small tick sizes such as France-Telecom (FT), one finds a remarkably linear behavior for $\mathcal{D}(\ell)$, even for small ℓ. The absence of linear correlations in price changes is compatible with the idea that (statistical) arbitrage opportunies are absent, even for high frequency trading. In fact, in order to emphasize the differences from a strictly diffusive behavior, we have studied the quantity $\sqrt{\mathcal{D}(\ell)/\ell}$ (which has the dimension of Euros). We show this quantity in Fig. 3.10 for FT, averaged over three different periods: first semester of 2001 (where the tick size was 0.05 Euros), second semester of 2001, and the whole of 2002 (where the tick size was 0.01 Euros). One sees that $\mathcal{D}(\ell)/\ell$ is indeed nearly constant, with a small 'oscillation' on which we will comment later. Similar plots can be observed for other stocks (see Fig. 3.11). We have noted that for stocks with larger ticks, a slow decrease of $\mathcal{D}(\ell)/\ell$ is observed, corresponding to a slight anti-persistence (or sub-diffusion) effect. Note that since we study the mid-point, the anti-correlations reported here are not related to the trivial bid-ask bounce.

The conclusion is that the random walk (diffusive) behavior of stock prices appears even at the trade by trade level, with a diffusion constant D which is of the order of the typical bid-ask squared. From Fig. 3.10, one indeed sees that $\sqrt{\mathcal{D}(1)} \sim 0.01$ Euros, which is precisely the tick size, whereas FT has an average bid-ask spread equal to two ticks. Hence, each transaction typically moves the mid-point by half the bid-ask.

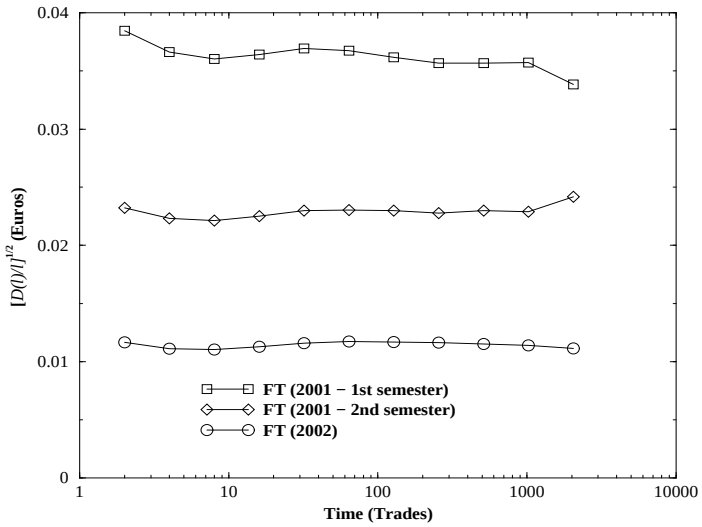

Fig. 3.10: Plot of $\sqrt{\mathcal{D}(\ell)/\ell}$ as a function of ℓ for France-Telecom, during three different periods. The variation of $\mathcal{D}(\ell)/\ell$ with ℓ is very small, in particular in the small tick (0.01 Euros) period (July 2001 – December 2002). For the large tick size period (0.05 Euros; January 2001 – June 2001), there is a systematic downward trend: see also Fig. 3.11.

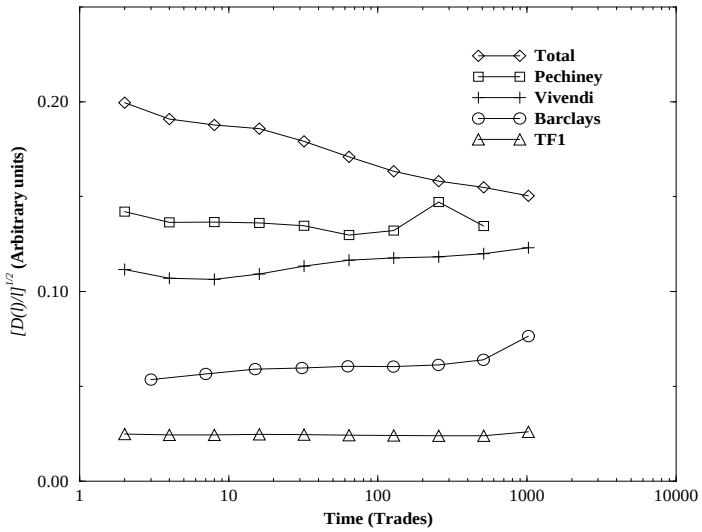

Fig. 3.11: Plot of $\sqrt{\mathcal{D}(\ell)/\ell}$ as a function of ℓ for other stocks during the year 2002, except Barclays (May-June 2002). The y-axis has been rescaled arbitrarily for clarity. We note that stocks with larger tick size tend to reveal a stronger mean-reverting effect.

Response function and market impact

In order to better understand the impact of trading on price changes, one can study the following *response function* $\mathcal{R}(\ell)$, defined as:

$$\mathcal{R}(\ell) = \langle (p_{n+\ell} - p_n) \cdot \varepsilon_n \rangle \,, \tag{3.24}$$

where ε_n is the sign of the n-th trade. The quantity $\mathcal{R}(\ell)$ measures how much, on average, the price moves up conditioned to a buy order at time 0 (or a sell order moves the price down) a time ℓ later. As will be clear below, this quantity is however not the market response to a single trade, a quantity that will later be denoted by G_0. A more detailed object can in fact be defined by conditioning the average to a certain volume V of the n-th trade:

$$\mathcal{R}(\ell, V) = \langle (p_{n+\ell} - p_n) \cdot \varepsilon_n \rangle|_{V_n = V} \,. \tag{3.25}$$

Previous empirical studies have mostly focused on the volume dependence of $\mathcal{R}(\ell, V)$, and established that this function is strongly concave as a function of the volume (Hasbrouck, 1991; Barra, 1997; Kempf & Korn, 1998; Plerou et al., 2002, Hopman, 2002. In (Lillo et al., 2003), a thorough analysis of U.S. stocks was analyzed in terms of a piecewise power-law dependence for $\mathcal{R}(\ell = 1, V) \propto V^{\alpha}$, with an exponent $\alpha \simeq 0.4$ for small volumes, and a smaller value ($\alpha \simeq 0.2$) for larger volumes; see also (Lillo & Farmer, 2003). It has been proposed (Potters & Bouchaud, 2003) that this dependence might in fact be logarithmic (see also a footnote in (Plerou et al., 2002)): $\mathcal{R}(\ell = 1, V) = R_1 \ln V$ (where R_1 is a stock dependent constant), a law that seems to satisfactorily account for all the data that we have analyzed. The empirical determination of the temporal structure of $\mathcal{R}(\ell, V)$ has been much less investigated (although one can find in (Plerou et al., 2002) somewhat related results on a coarse-grained version of $\mathcal{R}(\ell, V)$). $\mathcal{R}(\ell, V)$ can be written in a factorized form (Potters & Bouchaud, 2003) (first suggested on theoretical grounds in (Daniels et al., 2001)):

$$\mathcal{R}(\ell, V) \approx \mathcal{R}(\ell) f(V); \qquad f(V) \propto \ln V, \tag{3.26}$$

where $\mathcal{R}(\ell)$ is a slowly varying function that initially increases up to $\ell \sim 100 - 1000$ and then is seen to decrease back, with a rather small overall range of variation. The initial increase of $R(\ell)$ was reported in (Hasbrouck, 1991) and has also recently been noticed by Lillo and Farmer (Lillo & Farmer, 2003). Here, we provide much better data that supports both the above assertions. We show for example in Fig. 3.12 the temporal structure of $\mathcal{R}(\ell)$ for France Telecom, for different periods. Note that $\mathcal{R}(\ell)$ increases by a factor ~ 2 between $\ell = 1$ and $\ell = \ell^* \approx 1000$, before decreasing back. Similar

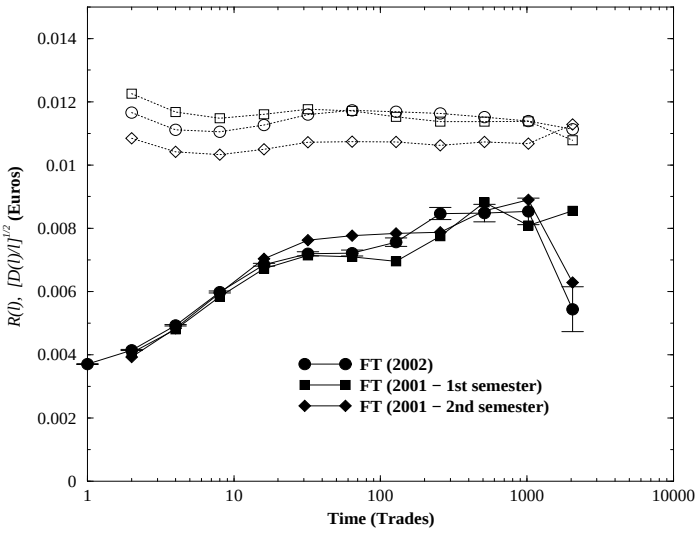

Fig. 3.12: Average response function $\mathcal{R}(\ell)$ for FT, during three different periods (black symbols). We have given error bars for the 2002 data. For the 2001 data, the $y-$axis has been rescaled to best collapse onto the 2002 data. Using the same rescaling factor, we have also shown the data of Fig. 1. The fact that the same rescaling works approximately for $\mathcal{D}(\ell)$ as well will be dwelled on further in section 3.3.1 below.

results have been obtained for many different stocks as well: Fig. 3.13 shows a small selection of other stocks, where the non monotonous behavior of $\mathcal{R}(\ell)$ is shown. However, in some cases (such as Pechiney), the maximum is not observed. One possible reason is that the number of daily trades is in this case much smaller (~ 1000), and that ℓ^* is beyond the maximum intra-day time lag. On the other hand, the model discussed below does also allow for monotonous response functions.

The existence of a time scale ℓ^* beyond which $\mathcal{R}(\ell)$ decreases is thus both statistically significant, and to a large degree independent of the considered stock. On the other hand, the amplitude of the change of $\mathcal{R}(\ell)$ seems to be stock dependent. As will be clear later, the fact that $\mathcal{R}(\ell)$ slowly increases before decreasing back to negative values is a non trivial result that requires a specific interpretation.

Turning now to the factorization property of $\mathcal{R}(\ell, V)$, Eq. (3.26), we illustrate its validity in Fig. 3.14, where $\mathcal{R}(\ell, V)/f(V)$ is plotted as a function of ℓ for different values of V. The function $f(V)$ was chosen for best visual rescaling, and is found to be close to $f(V) = \ln V$, as expected. Note that for the smallest volume (open circles), the long time behavior of $\mathcal{R}(\ell, V)$ seems

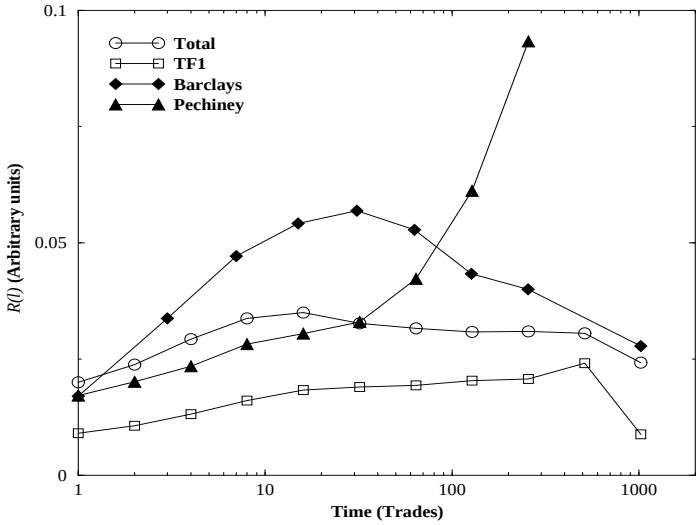

Fig. 3.13: Average response function $\mathcal{R}(\ell)$ for a restricted selection of stocks, during the year 2002.

to be different, which is probably due to the fact that small volumes are in fact more likely to be large volumes chopped up into small pieces.

Fraction of informed trades

One has to keep in mind that the response function $\mathcal{R}(\ell)$ captures a *small*, but systematic effect that relates the average price change to the sign of a trade. The fluctuations around this small signal are large, and increase with ℓ. A way to see this is to introduce the random variable $u_\ell = (p_{n+\ell} - p_n).\varepsilon_n$. By definition, $\mathcal{R}(\ell)$ is the average of u_ℓ, and $\mathcal{D}(\ell)$ is the average of u_ℓ^2. Since $\mathcal{R}(\ell)$ is roughly constant whereas $\mathcal{D}(\ell)$ grows linearly with ℓ, one sees that the impact of a given trade (as measured by $\mathcal{R}(\ell)$) rapidly becomes lost in the fluctuations.

In Fig. 3.15, we show the whole empirical distribution $P(u_\ell)$ of u_ℓ for $\ell = 128$ (other values of ℓ will be discussed below). This distribution is found to be only slightly skewed in the direction of positive u_ℓ. In fact, if one considers the shifted variable $u_\ell - \nu$, where $\nu = 0.01$ Euros, the distribution becomes nearly symmetric. Note that 0.01 Euros is equal to half the typical bid-ask spread and can therefore be seen as the minimal cost of a market order.

Empirical price changes are known to be highly kurtic, with sudden jumps separated by less volatile, random walk like periods. Market orders, that incur an immediate cost, are usually interpreted as 'informed' trades. This pre-

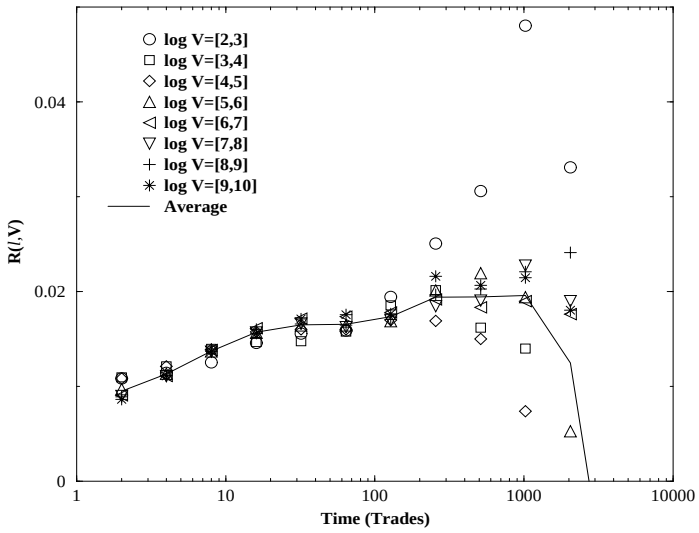

Fig. 3.14: Average response function $\mathcal{R}(\ell, V)$, conditioned to a certain volume V, as a function of ℓ. Data for different V's have been divided by $f(V) \propto \ln V$ such as to obtain good data collapse. The thick line corresponds to $\mathcal{R}(\ell)$ (unscaled).

sumably means that these trades correctly anticipate significant price changes as a result of some private information, that justifies paying a liquidity cost for immediate execution. Therefore, in this picture, the non zero value of $\langle u_\ell \rangle$ should mostly be due to the fraction of informed trades. Noise induced trades, on the other hand, should only temporarily move the price, and be uncorrelated with the future value of the stock. Hence, the positive tail of the distribution $P(u_\ell)$ (corresponding to informed trades) should be significantly fatter than the negative tail.

Now, the nearly symmetric shape of $P(u_\ell - \nu)$ in Fig. 3.15 shows that one can hardly detect the statistical presence of such trades that correctly anticipate the sign of the price change on a short term basis, such as to cover at least minimal trading costs.

A Fluctuation-Response relation

In the study of Brownian particles, a very important result that dates back to Einstein relates the diffusion coefficient D to the response of the particle to an external force. That a similar relation might also hold in financial markets was first suggested by Rosenow (Rosenow, 2002), and substantiated there by some empirical results. We have performed an analysis related to, but different from that of Rosenow. For any given trading day, one can

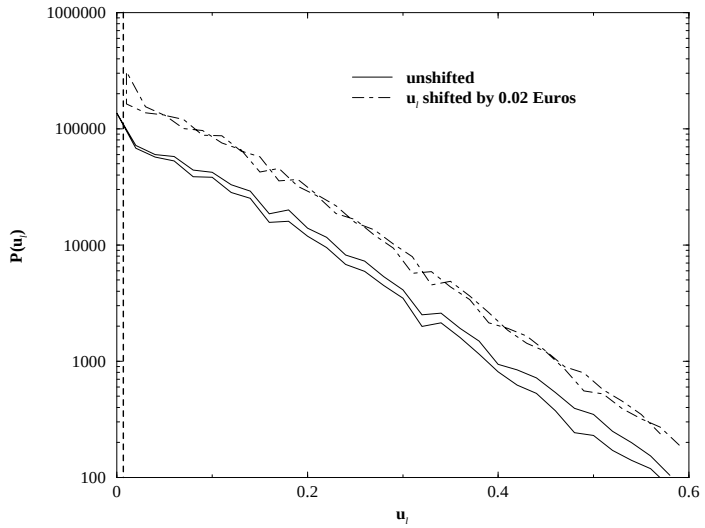

Fig. 3.15: Probability distribution $P(u_\ell)$ of the quantity $u_\ell = (p_{n+\ell} - p_n).\varepsilon_n$ (in Euros), for $\ell = 128$. The data is again FT during 2002. The negative part of the distribution has been folded back to positive u_ℓ in order to highlight the small positive skew of the distribution (which is seen to increase slightly with $|u_\ell|$). The average value $\mathcal{R}(\ell) = \langle u_\ell \rangle$ is shown as the vertical dashed line. The dashed-dotted line corresponds to the distribution of $u_\ell - \nu$ with $\nu = 0.01$ Euros. This curve has been shifted upwards for clarity.

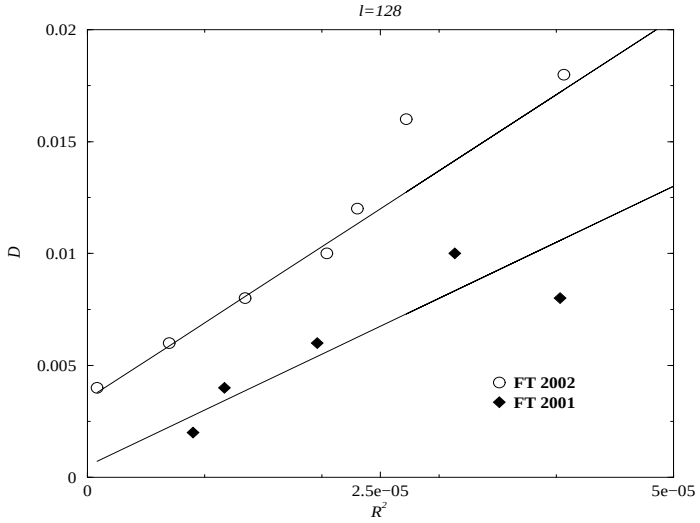

Fig. 3.16: Average diffusion constant $D = \mathcal{D}(\ell)/\ell$, computed for $\ell = 128$, and conditioned to a certain value of $\mathcal{R}^2(\ell)$, also computed for $\ell = 128$ (FT). The open symbols correspond to 2002, whereas the black symbols are computed using the first semester of 2001, where the tick size was 5 times larger. Correspondingly, the x-axis was rescaled down by a factor 25 and the y-axis by a factor five for this data set.

compute the average local diffusion constant $\mathcal{D}(\ell)$ over a given time scale, say $\ell = 128$, and the average local price response $\mathcal{R}(\ell)$ over the same time scale. Rosenow, on the other hand, computes a 'susceptibility' as the slope of the average price change over a given time interval versus the volume imbalance during the same time interval (see (Plerou et al., 2002)), and relates this susceptibility to the diffusion constant. The analogue of Rosenow's result (Rosenow, 2002) (which was motivated by a Langevin equation for price variations – see (Bouchaud & Cont, 1998)), is a linear relation between $\mathcal{R}^2(\ell)$ and $\mathcal{D}(\ell)$, which we illustrate in Fig. 3.17 for FT, for two different periods (first semester of 2001, and 2002). A similar result can also be read from Fig. 3.12. As will be clear in the following, such a relation will appear naturally within the simple model that we introduce in Section 3.

Long term correlation of trade signs

Here is our main empirical result. Although, as mentioned above, the statistics of price changes reveals very little temporal correlations, the correlation function of the sign ε_n of the trades, on the other hand, reveals very slowly

decaying correlations as a function of trade time. This correlation has been mentioned in some papers before, see e.g. (Hasbrouck, 1991; Hopman, 2002). Here, we propose that these correlations decay as a power-law of the time lag, at least up to $\ell \approx 15000$ (two trading days) beyond which we do not have sufficiently accurate data.

More precisely, one can consider the following correlation function:

$$\mathcal{C}_0(\ell) = \langle \varepsilon_{n+\ell}\varepsilon_n \rangle - \langle \varepsilon_n \rangle^2 \tag{3.27}$$

If trades were random, one should observe that $\mathcal{C}_0(\ell)$ decays to zero beyond a few trades. Surprisingly, this is not what happens: on the contrary, $\mathcal{C}_0(\ell)$ is strong and decays very slowly toward zero, as an inverse power-law of ℓ (see Fig. 9):

$$\mathcal{C}_0(\ell) \simeq \frac{C_0}{\ell^\gamma}, \qquad (\ell \geq 1). \tag{3.28}$$

The value of γ seems to be somewhat stock dependent. For example, for FT, one finds $\gamma \approx 1/5$, whereas for Total $\gamma \approx 2/3$. In their study, Lillo and Farmer found a somewhat larger value of $\gamma \approx 0.39$ for Vodafone (Lillo & Farmer, 2003). In any case, the value of γ is found to be smaller than one, which is very important because the integral of $\mathcal{C}_0(\ell)$ is then *divergent*. This is in fact the precise definition of 'long-term' correlations. Now, as will be shown more precisely in the next section, the integral of $\mathcal{C}_0(\ell)$ can intuitively be thought of as the effective number N_e of correlated successive trades. Hence, out of – say – 1000 trades, one should group together

$$N_e \simeq 1 + \sum_{\ell=1}^{1000} \mathcal{C}_0(\ell) \approx 1 + \frac{C_0}{1-\gamma} 1000^{1-\gamma} \tag{3.29}$$

'coherent' trades. For FT, $\gamma \approx 1/5$ and $C_0 \approx 0.2$, which means that the effect of one trade should be amplified, through the correlations, by a factor $N_e \approx 50$! In other words, both the response function $\mathcal{R}$ and the diffusion constant should increase by a factor 50 between $\ell = 1$ and $\ell = 1000$, in stark contrast with the observed empirical data. This is the main puzzle that one should try to elucidate: how can one reconcile the strong, slowly decaying correlations in the sign of the trades with the nearly diffusive nature of the price fluctuations, and the nearly structureless response function?

Before presenting a mathematical transcription of the above question and proposing a possible resolution, let us comment on two related correlation functions that will naturally appear in the following, namely:

$$\mathcal{C}_1(\ell) = \langle \varepsilon_{n+\ell}\ \varepsilon_n \ln V_n \rangle, \tag{3.30}$$

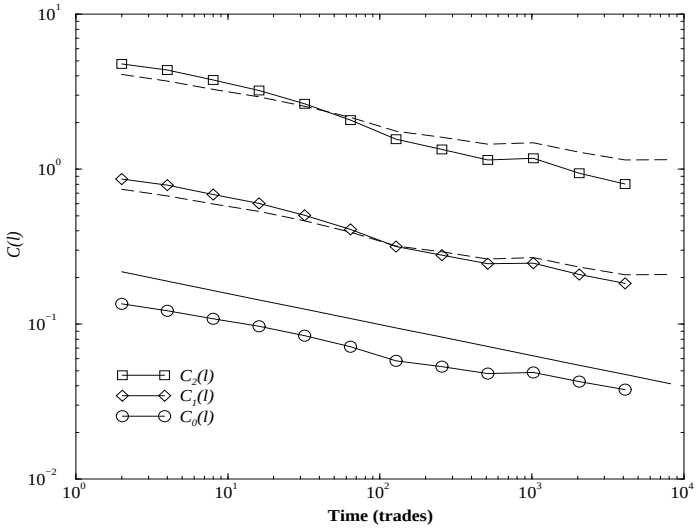

Fig. 3.17: Volume weighted sign autocorrelation functions as a function of time lag: $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$ (see text for definitions). The straight line corresponds to $\ell^{-\gamma}$ with $\gamma = 1/5$. The dotted lines correspond to the simple approximation given by Eqs. (3.32). Using data across different days allows one to extent the power-law dacay at least up to 15000 trades.

and

$$\mathcal{C}_2(\ell) = \langle \varepsilon_{n+\ell} \ln V_{n+\ell} \; \varepsilon_n \ln V_n \rangle. \tag{3.31}$$

We have found empirically that these two 'mixed' correlation functions are proportional to $\mathcal{C}_0(\ell)$ (see Fig 8):

$$\mathcal{C}_1(\ell) \approx \langle \ln V \rangle \mathcal{C}_0(\ell); \qquad \mathcal{C}_2(\ell) \approx \langle \ln V \rangle^2 \mathcal{C}_0(\ell). \tag{3.32}$$

There are however small systematic deviations, which indicate that (i) small volumes contribute more to the long range correlations that larger volumes and (ii) $\ln V - \langle \ln V \rangle$ is a quantity exhibiting long range correlations as well.

3.3.2 A micro-model of price fluctuations

Set up of the model

In order to understand the above results, we will postulate the following trade superposition model, where the price at time n is written as a sum over all

past trades, of the impact of one given trade propagated up to time n: [3]

$$p_n = \sum_{n'<n} G_0(n-n')\,\varepsilon_{n'} \ln V_{n'} + \sum_{n'<n} \eta_{n'} + \epsilon_n, \tag{3.33}$$

where $G_0(.)$ is the 'bare' impact function (or propagator) of a single trade, that we assume to be a fixed, non random function that only depends on time differences (which is a rather strong assumption, see (Lillo & Farmer, 2003)). The η_n are also random variables, assumed to be independent from the ε_n and model all sources of price changes not described by the direct impact of the trades: the bid-ask can change as the result of some news, or of some order flow, in the absence of any trades. We will in the following assume that the η_n are also uncorrelated in time, although this assumption can easily be relaxed. Finally, the ϵ_n are independent (zero mean) random variables that can be seen as a high frequency noise on the position of the bid and of the ask, inducing some noise on the determination of the midpoint.

The bare impact function $G_0(\ell)$ represents by definition the average impact of a single trade after ℓ trades. It could be in principle measured empirically by launching on the market a sequence of real trades of totally random signs, and averaging the impact over this sample of trades (a potentially costly experiment!).[4] As will be clear below, the difference between the quantity $\mathcal{R}(\ell)$ introduced in the previous Section and $G_0(\ell)$ in fact comes from the strong autocorrelation of the sign of the trades. In order to understand the temporal structure of $G_0(\ell)$, note that a single trade first impacts the midpoint by changing the bid (or the ask). But then the subsequent limit order flow due to that particular trade might either center on average around the new midpoint (in which case $G_0(\ell)$ would be constant), or, as we will argue below, tend to mean revert toward the previous midpoint (in which case $G_0(\ell)$ decays with ℓ).

Using this representation, the price increment between an arbitrarily chosen initial time 0 and time ℓ is:

$$\begin{aligned} p_\ell - p_0 &= \sum_{0\le n<\ell} G_0(\ell-n)\varepsilon_n \ln V_n + \sum_{n<0} \left[G_0(\ell-n) - G_0(-n)\right] \varepsilon_n \ln V_n \\ &+ \sum_{0\le n<\ell} \eta_n + \epsilon_\ell - \epsilon_0. \end{aligned} \tag{3.34}$$

If the signs ε_n were independent random variables, both the response function and the diffusion would be very easy to compute. For example, one would

[3] A similar model, but linear in V_n and truncated beyond only a few trades in the past, was discussed in (Hasbrouck, 1991).

[4] However, such an experimental protocol might induce 'copy-cat' trades and still lead to a difference between the measured response function and G_0

have:[5]

$$\mathcal{R}_t(\ell) = \langle \ln V \rangle G_0(\ell), \tag{3.35}$$

i.e. the observed impact function and the bare response function would be proportional. Similarly, one would have:

$$\mathcal{D}_t(\ell) = \langle \ln^2 V \rangle \left(\sum_{0<n\leq\ell} G_0^2(n) + \sum_{n>0} \left[G_0(\ell+n) - G_0(n) \right]^2 \right) + D_\eta \ell + 2D_0, \tag{3.36}$$

where D_η is the variance of the η's and D_0 is the variance of the ϵ's. In the simplest case of a constant bare impact function, $G_0(\ell) = \Gamma_0$ for all $\ell > 0$, one then finds (for $D_0 = 0$) a pure diffusive behavior, as expected:

$$\mathcal{D}_t(\ell) = \ell \left[\langle \ln^2 V \rangle \Gamma_0^2 + D_\eta \right]. \tag{3.37}$$

This case (no correlations between the ε's and a constant bare impact function) corresponds to the simplest possible zero intelligence market, where agents are memoryless. However, we have seen that in fact the ε's have long range correlations. In this case, the average response function reads:

$$\mathcal{R}_t(\ell) = \langle \ln V \rangle G_0(\ell) + \sum_{0<n<\ell} G_0(\ell-n) \mathcal{C}_1(n) + \sum_{n>0} \left[G_0(\ell+n) - G_0(n) \right] \mathcal{C}_1(n). \tag{3.38}$$

Note in passing that our trade superposition model, Eq. (3.33), together with Eq. (3.32) leads to the factorization property mentioned above (see Fig. 5):

$$\mathcal{R}_t(\ell, V) = \frac{\ln V}{\langle \ln V \rangle} \mathcal{R}_t(\ell). \tag{3.39}$$

Now, one sees more formally the paradox discussed in the previous Section: assuming that the impact of each trade is permanent, i.e. $G_0(\ell) = \Gamma_0$, leads to:

$$\mathcal{R}_t(\ell) = \Gamma_0 \left[\langle \ln V \rangle + \sum_{0<n<\ell} \mathcal{C}_1(n) \right]. \tag{3.40}$$

If $\mathcal{C}_1(n)$ decays as a power-law with an exponent $\gamma < 1$, then the average impact $\mathcal{R}(\ell)$ should *grow* like $\ell^{1-\gamma}$, and therefore be amplified by a very large factor as ℓ increases, at variance with empirical data. The only way out of this conundrum is (within the proposed model) that the bare impact function $G_0(\ell)$ itself should decay with time, in such a way to offset the amplification effect due to the trade correlations.

[5] In the following, we will use the subscript 't' to denote the theoretical expressions for the response function or diffusion.

A relation between the bare propagator and the sign correlation function

In order to get some guidance, let us now look at the general formula for the diffusion. After a few lines of calculations, one finds:

$$\mathcal{D}_t(\ell) = \langle \ln^2 V \rangle \left[\sum_{0 \le n < \ell} G_0^2(\ell - n) + \sum_{n>0} \left[G_0(\ell + n) - G_0(n) \right]^2 \right] \quad (3.41)$$

$$+ \quad 2\Delta(\ell) + D_\eta \ell + 2D_0, \quad (3.42)$$

where $\Delta(\ell)$ is the correlation induced contribution:

$$\Delta(\ell) = \sum_{0 \le n < n' < \ell} G_0(\ell - n) G_0(\ell - n') \mathcal{C}_2(n' - n) \quad (3.43)$$

$$+ \sum_{0 < n < n'} \left[G_0(\ell + n) - G_0(n) \right] \left[G_0(\ell + n') - G_0(n') \right] \mathcal{C}_2(n' - n)$$

$$+ \sum_{0 \le n < \ell} \sum_{n' > 0} G_0(\ell - n) \left[G_0(\ell + n') - G_0(n') \right] \mathcal{C}_2(n' + n). \quad (3.44)$$

The constraint from empirical data is that this expression must be approximately linear in ℓ. As shown in the section 3.3.3, the requirement that $\mathcal{D}_t(\ell)$ is strictly linear in ℓ for all ℓ in fact allows one to express $G_0(\ell)$ as a function of $\mathcal{C}_2(\ell)$. Here, we present a simple asymptotic argument. If we make the ansatz that the bare impact function $G_0(\ell)$ also decays as a power-law:

$$G_0(\ell) = \frac{\Gamma_0 \ell_0^\beta}{(\ell_0 + \ell)^\beta} \qquad (\ell \ge 1) \quad (3.45)$$

then one can estimate $\mathcal{D}_t(\ell)$ in the large ℓ limit. When $\gamma < 1$, one again finds that the correlation induced term $\Delta(\ell)$ is dominant, and all three terms scale a $\ell^{2-2\beta-\gamma}$, provided $\beta < 1$. In other words, the Hurst exponent of price changes is given by $2H = 2 - 2\beta - \gamma$. Therefore, the condition that the fluctuations are diffusive at long times ($H = 1/2$) imposes a relation between the decay of the sign autocorrelation γ and the decay of the bare impact function β that reads:

$$2\beta + \gamma = 1 \longrightarrow \beta_c = \frac{1 - \gamma}{2} \quad (3.46)$$

For $\beta > \beta_c$, the price is *sub-diffusive* ($H < 1/2$), which means that price changes show anti-persistence; while for $\beta < \beta_c$, the price is *super-diffusive* ($H > 1/2$), i.e. price changes are persistent. For FT, $\gamma \approx 1/5$ and therefore $\beta_c \approx 2/5$.

As shown in the section 3.3.3 , one can in fact obtain an exact relation between $G_0(\ell)$ and $\mathcal{C}_2(\ell)$ if one assumes that price changes are strictly uncorrelated (i.e. that $\mathcal{D}(\ell)$ is linear in ℓ for all ℓ). The asymptotic analysis of this relation leads, not surprisingly, to the same exponent relation $\beta_c = (1-\gamma)/2$ as above.

At this stage, there seems still to be a contradiction with empirical data, for if one goes back to the response function given by Eq. (3.38), one finds that whenever $\beta+\gamma < 1$ (which is indeed the case for $\beta = \beta_c$ and $\gamma < 1$), the dominant contribution to $\mathcal{R}_t(\ell)$ should behave as $\ell^{1-\beta-\gamma}$ and thus *grow* with ℓ. For example, for $\gamma \approx 1/5$ and $\beta \approx 2/5$, one should find that $\mathcal{R}_t(\ell) \propto \ell^{2/5}$, which is incompatible with the empirical data of Figs. 3.12 and 3.13. But the surprise comes from the numerical prefactor of this power law. One finds, for large ℓ:

$$\mathcal{R}_t(\ell) \simeq \langle \ln V \rangle \Gamma_0 C_0 \frac{\Gamma(1-\gamma)}{\Gamma(\beta)\Gamma(2-\beta-\gamma)} \left[\frac{\pi}{\sin \pi\beta} - \frac{\pi}{\sin \pi(1-\beta-\gamma)} \right] \ell^{1-\beta-\gamma}. \tag{3.47}$$

Therefore, only when $\beta = \beta_c$, is the prefactor exactly zero, and leads to the possibility of a nearly constant impact function! For faster decaying impact functions (larger β's), this prefactor is negative, whereas for more slowly decaying impact functions this prefactor is positive.[6] Interestingly, even if the bare response function $G_0(\ell)$ is positive for all ℓ, the average response $\mathcal{R}_t(\ell)$ can become negative for large enough β's, as a consequence of the correlations between trades.

Fitting the average response function

Since the dominant term is zero for the 'critical' case $\beta = \beta_c$, and since we are interested in the whole function $\mathcal{R}_t(\ell)$ (including the small ℓ regime), we have computed $\mathcal{R}_t(\ell)$ numerically, by performing the discrete sum Eq. (3.38) exactly, and fitted it to the empirical response $\mathcal{R}$. The results are shown in Fig. 3.18. We have fixed the parameters γ and C_0 to the values extracted from the behavior of $\mathcal{C}_1(\ell)$ (see Fig. 8): $\gamma = 0.24$ and $C_0 = 0.20$. The overall scaling parameter Γ_0 is adjusted to match the value of $\mathcal{R}(\ell = 1)$.[7] The values of β and ℓ_0 are fitting parameters: we show in Fig. 3.18 the response function computed for different values of β in the vicinity of $\beta_c = 0.38$, and used $\ell_0 = 20$.

[6] Note that although this prefactor increases (in absolute value) with β for $\beta > \beta_c$, the power of ℓ decreases, which means that for large ℓ the amplitude of $\mathcal{R}_t(\ell)$ decreases with β, as intuitively expected.

[7] The numerical value of Γ_0 is found to be such that: $\Gamma_0 \ell_0^\beta = 2.8\,10^{-3}$ Euros.

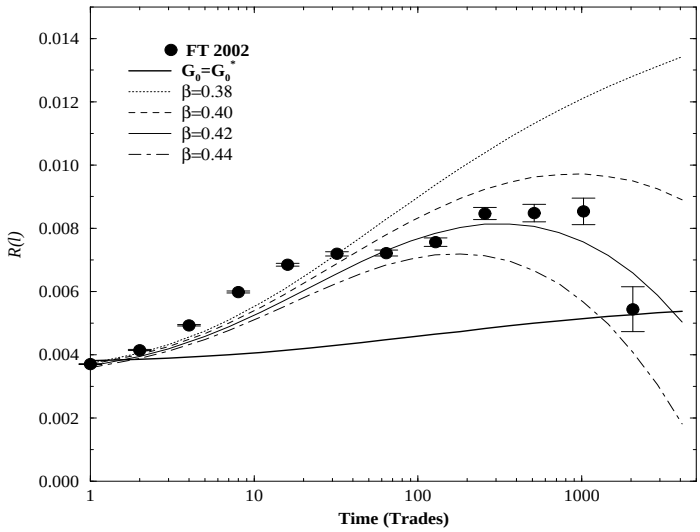

Fig. 3.18: Theoretical impact function $\mathcal{R}_t(\ell)$, from Eq. (3.38), and for different values of β close to $\beta_c = 0.38$. The shape of the empirical response function can be quite accurately reproduced using $\beta = 0.42$. The only remaining free parameter is $\ell_0 = 20$. The thick plain line is $\mathcal{R}_t(\ell)$ computed using the 'pure diffusion' propagator G_0^* determined in the section 3.3.3, Eq. (3.54).

The results are compared with the empirical data for FT, showing that one can indeed satisfactorily reproduce, when $\beta \approx \beta_c$, a weakly increasing impact function that reaches a maximum and then decays. One also sees, from Fig. 3.18, that the relation between β and γ must be quite accurately satisfied, otherwise the response function shows a distinct upward trend (for $\beta < \beta_c$) or a downward trend ($\beta > \beta_c$).[8] We have tried other simple forms for $G_0(\ell)$, such as a simple exponential decay toward a possibly non zero asymptotic value, but this leads to unacceptable shapes for $\mathcal{R}(\ell)$. Of course, a more precise fit of the initial increase of $\mathcal{R}(\ell)$ seen in Fig. 3.18 could be achieved by choosing a more complicated function $G_0(\ell)$, that first increases slightly before decaying to zero.

It is also interesting to use the propagator G_0^* determined in the section 3.3.3 from the assumption of a purely diffusive price process for all ℓ's. This propagator is plotted in Fig. 3.19, and compared to the G_0 determined above from the fit of $\mathcal{R}(\ell)$. As shown in Fig. 3.18, the use of G_0^* does *not* lead to a very good fit of $\mathcal{R}(\ell)$. Since the latter quantity is in fact *very sensitive to the chosen shape* for G_0, it does reveal small, but systematic deviations

[8] The former scenario might actually explain the different behavior of Pechiney seen in Fig.3.13.

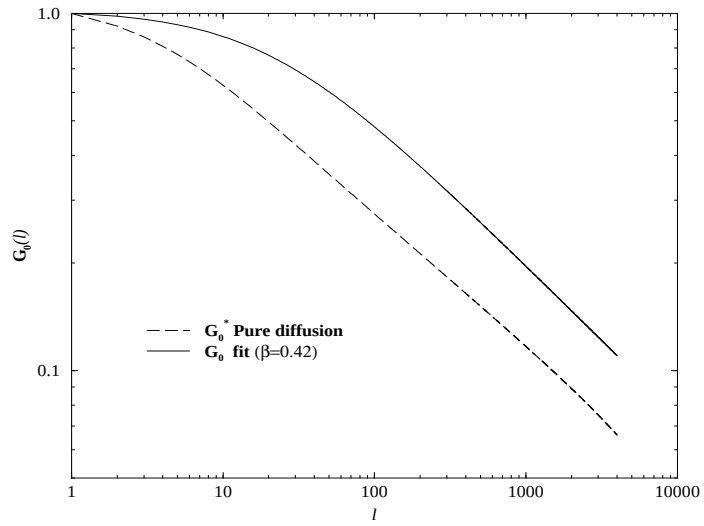

Fig. 3.19: Shape of the bare propagator G_0, determined either by the fit of $\mathcal{R}$, with $\beta = 0.42$ and $\ell_0 = 20$, or using the exact relation, Eq. (3.54), derived in the subsection 4 from the assumption of a purely diffusive process.

from a purely diffusive price process. [Note that if one had $\mathcal{C}_2(\ell) = \mathcal{C}_1(\ell)$, the resulting $\mathcal{R}(\ell)$ should be strictly constant.]

Back to the diffusion constant

As we showed above, the reason for the fine tuning of β is the requirement that price changes are almost diffusive. We can therefore also compute $\mathcal{D}_t(\ell)$ for all values of ℓ using the very same values of γ, β, C_0, ℓ_0 and Γ_0. Now, in order to fit the data one has two extra free parameters: D_η and D_0 (see Eq. (3.41)). With these two extra parameters, one can reproduce the empirical determination of $\mathcal{D}(\ell)/\ell$ (see Fig. 3.20). The small deviations of this quantity from a horizontal line at finite ℓ are due to the difference between G_0 and G_0^* and/or to the possible autocorrelations between the η_n variables, which we have neglected here. Note that the contribution of the term D_η turns out to be a factor two larger than that of the impact contribution, Eq. (3.41), which means that the small increase of the 'impact contribution' with ℓ (lower graph of Fig. 3.19) is hardly detectable in $\mathcal{D}(\ell)/\ell$.

Coming back to the Fluctuation-Response relation discussed in section 3.3.1, we see that our model predicts, for $\ell \gg 1$ where the effect of D_0 can

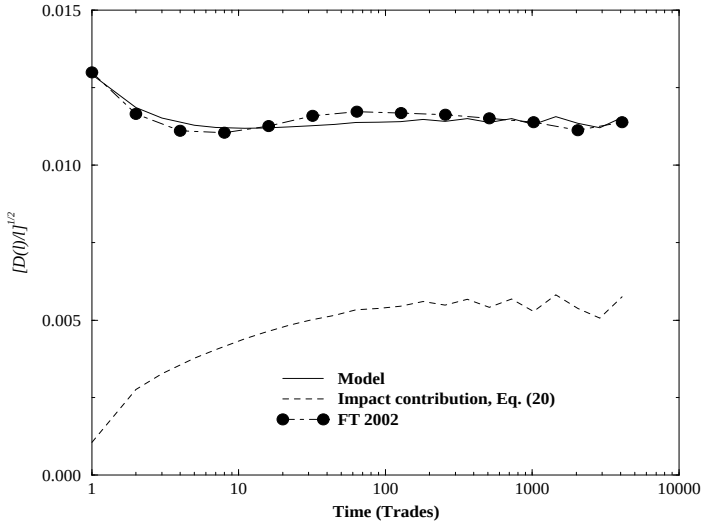

Fig. 3.20: Diffusion constant $\mathcal{D}(\ell)/\ell$, using Eq. (3.41), with the values of γ, β, C_0, ℓ_0 and Γ_0 determined from $\mathcal{R}(\ell)$. Two extra parameters were used: $D_\eta = 10^{-4}$ and $D_0 = 3.3\,10^{-5}$ (both in Euro squared). The lower graph is the 'impact contribution' to $\mathcal{D}_t(\ell)$, given by Eq. (3.41) with $D_\eta = 0$. The 'oscillations' at long times is a numerical artefact.

be neglected:

$$\frac{\mathcal{D}_t(\ell)}{\ell} = Z\langle \ln V\rangle^2 C_0 \Gamma_0^2 + D_\eta, \qquad \mathcal{R}_t(\ell) = Z'\langle \ln V\rangle \Gamma_0 C_0, \tag{3.48}$$

where Z, Z' are numerical constants. Assuming that from one day to the next both the average (log-)traded volume and the impact Γ_0 of each individual trade might change, while C_0 is fixed, immediately leads to the affine relation between $\mathcal{D}$ and $\mathcal{R}^2$ reported in section 3.3.1.

Discussion

The conclusion of this Section is that our 'micro-model' of prices, Eq. (3.33), can be used as a theoretical canvas to rationalize and interpret the empirical results found in the previous Section. Most surprising is the constraint that the empirical results impose on the shape of the 'bare' response function G_0, which is found to be a slowly decaying power law which must precisely cancel the slowly decaying autocorrelation of the trades, but reveals systematic deviations from a pure diffusion process, hardly noticeable on the diffusion constant itself. The fact that the bare impact function decays with time (at least on intra-day time scales), in a finely tuned way to compensate

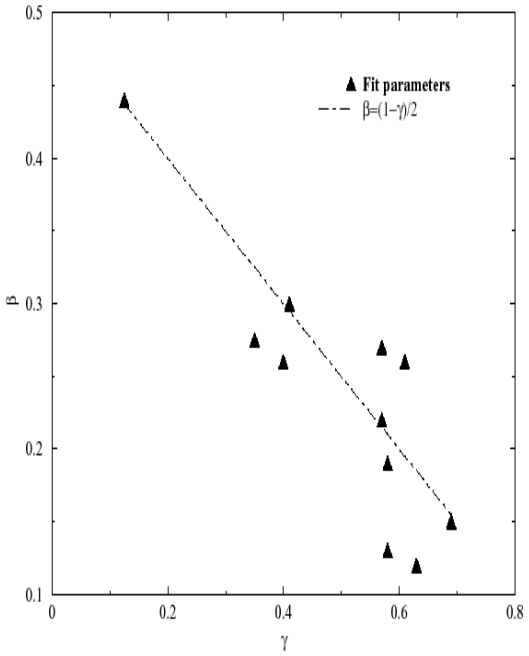

Fig. 3.21: Scatter plot of the exponents β, γ extracted from the fit of G_0 and C. These exponents are seen to lie in the vicinity of the critical line $\beta = (1-\gamma)/2$ (dotted line), as expected from the nearly diffusive behavior of prices.

the long memory in the trades, is the central result of this section. This effect is lost in the zero intelligence models of Poisonnian order flows, where, after decreasing during a short transient, the impact of each trade becomes permanent: $G_0(\ell) \to G_\infty > 0$. In fact, both the long time memory of the trades and the slowly relaxing impact function reported here must be the consequence of the strategic behavior of market participants, that we discuss below in order to get an intuitive understanding of the mechanisms at play.

Although our detailed analysis concerns FT, it is clear that our conclusions are more general, since both the strong autocorrelations in the trade signs, the near constancy of the average response function and the diffusive nature of price changes have been observed on all stocks, with only quantitative changes (see Figs. 3.11 and 3.13). Fig.(3.21), taken from (Bouchaud, Kockelkoren & Potters, 2004), shows the fitting of the exponent γ and β for other stocks: Credit Agricole, Danone, Carrefour, Alcatel, Bouygues, Vivendi, Total, France-Telecom, Societe Generale, LVMH, Vivendi Env.

Finally, it would be very interesting to know whether the bare response

function levels off to a finite value for large time lags; this will require more data to go beyond the analysis of the present section to enlarge the available range of ℓ values. However, it seems reasonable to expect that $G_0(\ell)$ should indeed reach a finite asymptotic value for values of ℓ corresponding to a few days of trading.[9]

3.3.3 Critical balance of opposite forces: Market orders vs. limit orders

Stylized Market

Although trading occurs for a large variety of reasons, it is useful to recognize that traders organize in two broad categories:

- One is that of 'liquidity takers', that trigger trades by putting in market orders. The motivation for this category of traders might be to take advantage of some 'information', and make a profit from correctly anticipating future price changes. Information can in fact be of very different nature: fundamental (firm based), macro-economical, political, statistical (based on regularities of price patterns), etc. Since market orders allows one to be immediately executed, many impatient investors, who want to liquidate their position, or hedge, etc. might be tempted to place market orders, even at the expense of the bid-ask spread $S(t) = a(t) - b(t)$. Note that this definition is not strict, as agents willing to use private informations may also use limit orders.

- The other category is that of 'liquidity providers' (or 'market makers', although on electronic markets all participants can act as liquidity providers by putting in limit orders), who offer to buy or to sell but avoid taking any bare position on the market. Their profit comes from the bid-ask spread S: the sell price is always slightly larger than the buy price, so that each round turn operation leads to a profit equal to the spread S, at least if the midpoint has not changed in the mean time.

Long-memory in the market order flow

In order not to trigger a sudden increase of the ask a (or bid b) that would make their trade costly, liquidity takers obviously need to put on not too

[9] Hopman quotes three days as the time beyond which the autocorrelation of the trades sign falls to zero (Hopman, 2002), whereas we find that the power-law decay of this correlation persists up to at least two days of trading.

large orders. This is the rationale for dividing one's order in small chunks and disperse these as much as possible over time so as not to appear on the 'radar screens'. Doing so liquidity takers *necessarily create some temporal correlations* in the sign of the trades. Since these traders probably have a somewhat broad spectrum of volumes to trade (Gabaix et al., 2003), and therefore of trading horizons (from a few minutes to several weeks), this can explain the slow, power-law decay of the sign correlation function $\mathcal{C}_0(\ell)$ reported above.

Antagonist Forces

Our hypothesis is the following: when a chunk of market orders of same sign appears in the market order flow, limit orders of opposite sign are emitted. This effect reduces the impact of the correlated market orders. Formally, this leads to a decaying bare function, whose decay corresponds to the impact of the limit orders emitted in response to each market order. This description is supported by recent empirical data. In (Bouchaud, Kockelkoren, Potters, 2004) it is observed that the impacts of limits orders are anti-correlated with the market orders signs. Weber et al. (Weber, Rosenow, 2004) show empirically that the future limit order flow is anti-correlated with the price change: if the price goes up, more limit orders to sell are emitted. Their data is shown in Fig.(3.22).

This antagonist force may have several causes. Fig.(3.22) suggests the presence of a "latent" order-book: only after the price starts going up, some agents who were willing to sell emit limit orders. Acting this way they place limit orders little by little, and avoid the risk and the costs associated with large limit orders.

Such resting force may also be induced by the presence of market makers (Bouchaud, Gefen, Potters, Wyart, 2003), who have, by definition, a contrarian role. As we shall see in the next section, their gains diminish with volatility, and with the presence of trends in the price dynamics. Therefore they may be also be responsible for the opposition between market orders and limit order flows.

A striking aspect of these two antagonist forces is that they exactly compensate. This is true at the level of the exponents: the exponent of the bare impact is such that the price dynamics is diffusive ($H = 1/2$). More surprisingly, the compensation occurs also at the coefficients level, as $\frac{\mathcal{D}(\ell)}{\ell}$ is extremely flat, see Fig.(3.10). This is a necessary condition imposed by the absence of obvious statistical arbitrages: if the price was not perfectly diffusive, information would be contained in the past price dynamics, that could be used. In the following paragraph, inspired by a remark of Xavier Gabaix,

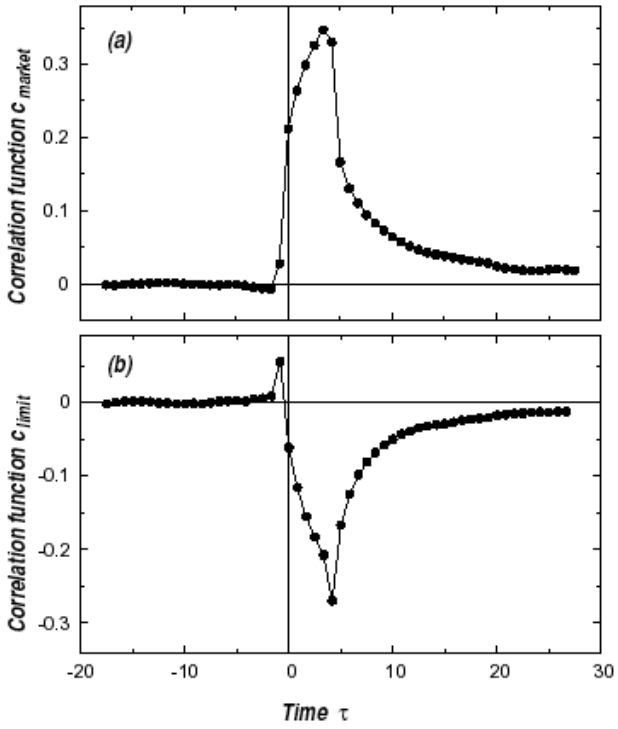

Fig. 3.22: Correlation functions between return (relative price change) and signed flow (buy minus sell orders). (a) Market orders and returns show strong positive equal time correlations decaying slowly to zero. (b) Limit orders preceding returns have weak positive correlations with them, while equal time correlations are strongly negative.

we show how such an arbitrage could take place. Imposing that all past information of the market order flow is implemented by the liquidity providers in the price leads to a perfectly diffusive behavior, and to the appearance of a bare impact function purely transient (if the correlation of market orders has no cut-off) with the correct exponent.

Model of perfect diffusion

Since the market order flow is correlated, each market order can be decomposed in two parts: a predictable part, that can be anticipated from the past market orders, and an unpredictable part, which represents the "new" information contained in the order flow. Here we show that if the liquidity providers change the price only according to the unpredictable part of the market order, then the price is perfectly diffusive *at all times* (rather than only asymptotically). This leads to an expression between the response to a single trade, the bare propagator G_0, and the sign correlation function $\mathcal{C}_2(\ell)$. In order to show this, let us assume that the random variable $q_n \equiv \varepsilon_n \ln V_n$ can be written as:

$$q_n = \sum_{m \leq n} K(n-m)\xi_m, \tag{3.49}$$

where ξ_n are uncorrelated random variables ($\langle \xi_n \xi_m \rangle = \langle \ln^2 V \rangle \delta_{n,m}$), and $K(.)$ a certain kernel. In order for the q_n to have the required correlations, the kernel $K(.)$ should obey the following equation:

$$\mathcal{C}_2(n) = \langle \ln^2 V \rangle \sum_{m \geq 0} K(m+n)K(m). \tag{3.50}$$

In the case where $\mathcal{C}_2$ decays as $\ell^{-\gamma}$ with $0 < \gamma < 1$, it is easy to show that the asymptotic decay of $K(n)$ should also be a power-law $n^{-\delta}$ with $2\delta - 1 = \gamma$. Note that $1/2 < \delta < 1$.

Inverting Eq. (3.49) allows one to obtain a set of uncorrelated random variables ξ_n from a set of correlated variables q_n:

$$\xi_n = \sum_{m \leq n} Q(n-m)q_m, \tag{3.51}$$

where Q is the matrix inverse of K, such that $\sum_{m=0}^{n} K(n-m)Q(m) = \delta_{m,n}$. Eqs. (3.49,3.51) in fact form the basis of linear filter theories, and ξ_n can be seen as the prediction error on the next variable q_n. ξ_n is the new information of the last market order, which cannot be deduced from the past.

Introducing discrete Laplace transforms:

$$\widehat{K}(E) = \sum_{n \geq 0} K(n)e^{-nE} \qquad \widehat{Q}(E) = \sum_{n \geq 0} Q(n)e^{-nE}, \tag{3.52}$$

one finds $\widehat{K}(E)\widehat{Q}(E) = 1$. For a power-law kernel $K(.)$, one obtains: $\widehat{Q}(E) \propto E^{1-\delta}$ for $E \to 0$, and therefore $Q(n) \propto n^{\delta-2}$ for large n. It is useful to note that in this case $\widehat{Q}(E=0) = \sum_{n\geq 0} Q(n) = 0$.

Now, it is clear that if agents only react to new informations, which is equivalent to define the price process p_n as:

$$p_n = \sum_{m<n} \xi_m, \tag{3.53}$$

then p_n is a diffusion process with a strictly linear $\mathcal{D}(\ell)$, since the ξ's are by construction uncorrelated. The price defined in this way can also be written, using Eq. (3.51), as a linear combination of past q_m's, as assumed in our micro-model Eq. (3.33), with:

$$G_0^*(\ell) \equiv \sum_{m=0}^{\ell-1} Q(m). \tag{3.54}$$

This is an exact relation between $\mathcal{C}_2$ (that allows one to compute in turn K and Q) and the response function G_0^* for all ℓ's, where the star indicates that strict diffusion is imposed.

In the case of power-law kernels, one finds from the above relation and from $Q(n) \propto n^{\delta-2}$ for large n:

$$G_0^*(\ell) \propto \ell^{\delta-1} \longrightarrow \beta = 1-\delta = \frac{1-\gamma}{2}, \tag{3.55}$$

which is, not surprisingly, the relation obtained in the main text from the assumption that prices are diffusive on long time scales.

3.3.4 Summary and Conclusion

The aim of this section was to study in details the statistics of price changes at the trade by trade level, and to analyze the interplay between the impact of each trade on the price and the volatility. Empirical data shows that (a) the price (midpoint) process is close to being purely diffusive, even at the trade by trade scale (b) the temporal structure of the impact function first increases and reaches a maximum after $100 - 1000$ trades, before decreasing back, with a rather limited overall variation (typically a factor 2) and (c) the sign of the trades shows surprisingly long range (power-law) correlations, at least up to 15000 trades (two trading days). The paradox is that if the impact of each trade was permanent, the price process should be strongly super-diffusive and the average response function should increase by a large factor as a function of the time-lag.

As a possible resolution of this paradox, we have proposed a micro-model of prices, Eq. (3.33) where the price at any instant is the causal result of all past trades, mediated by what we called a bare impact function, or propagator G_0, which describes the response to a single trade. All the empirical results can be reconciled if one assumes that this bare propagator also decays as a power-law in time, with an exponent which is precisely tuned to a critical value, ensuring simultaneously that prices are diffusive on long time scales and that the response function is nearly constant. Therefore, the seemingly trivial random walk behavior of price changes in fact results from a fined-tuned competition between two opposite effects, one leading to super-diffusion (the autocorrelation of trades) and the other leading to sub-diffusion (the decay of the bare impact function, induced by the emission of contrarian limit orders). The cancellation is however not exact: the non trivial behavior of the average response function allows one to detect small, but systematic deviations from a purely diffusive behavior, deviations that are hardly detectable on the price fluctuations themselves.

Then, we furnished an example of agents behavior that leads to a perfectly diffusive price. This takes place if the liquidity providers only change the price with respect to the unpredictable information contained in the market order flow. In this model too, the response to a single isolated trade, or bare propagator, decays in time. Such response is purely transient if the power law correlation of the market order signs does not display a cut-off at large times.

In what follows we shall use in particular Eq.(3.48), which indicates that despite markets order signs are correlated, a relation exists between the average square impact function and volatility.

3.4 Liquidity *vs* Volatility

The results presented in this section are preliminary [10]. Our main result is to show that if the average gain of limit orders and market orders are equal (and therefore zero)— as we expect in an electronic market where each agent can emit both types of order— then the spread is related to the average impact function. This relation has no fitting parameter, and is, as we shall see, in good agreement with empirical data. Furthermore, this relation does not rely on any extra assumptions on the agents behavior. In particular we do not need to assume any risk aversion from agents, nor a fraction of informed trades.

[10] We thank Julien Kockelkoren with whom this work is being done.

Then, using that the impact function is related to the volatility, as derived in Eq.(3.48), we relate volatility and spread. Using simple assumptions we show that this leads to the observed linear relation between volatility per trade and spread.

3.4.1 Market Making

In this section, we show that for a market making strategy to be profitable, there must be a relation between the spread and the response to market orders:

$$\langle SV \rangle_V \geq 2 \langle V \mathcal{R}(V) \rangle_V \tag{3.56}$$

where the average $\langle \rangle_V$ is made on the distribution of the trading volume. The term $\langle SV \rangle_V$ is not a priori equal to $\langle S \rangle_V \langle V \rangle_V$ since the market order size can be correlated with S. In a competitive environment, this relation must be an equality. If so, the costs associated with a limit order and a market order are equal. We expect that it is the case for a continuous double-auctions where any agent can emit both types of orders. This relation is valid even when the market order signs are correlated. It applies only for stocks where the spread is somewhat larger than the tick size.

We consider a market maker with a time horizon T (where T is in transaction time unit) who provides an infinitesimal fraction γ of the total available liquidity. The market maker puts limit orders both at the bid and at the ask. We assume that he participates to a fraction γ of every transaction volume [11]. His gains come from the spread. On the other hand, a market maker must have a balanced inventory at the end of some time horizon T. As we show now, this has a cost. We impose that at time T, the market maker must sell (or buy if he is short) the volume V that he has obtained during the trading period $[0, T]$, with $V = -\gamma \sum_{i=1}^{T} \epsilon_i V_i$, where V_i is the absolute value of the i^{th} market order. As γ is small, V is small too, and we may neglect the market maker impact, who trades back the volume V at the current price P_T. This is costly in general, because ΔP_T is anti-correlated with V. For example if the majority of the market orders were to buy, $\Delta P_T > 0$ and $V < 0$. As we shall see balancing this cost with the gain leads to Eq.(3.56).

When a market order of absolute size V_i is emitted at a time i, the mid-point price change at time $i + l$ is the random variable $(P_{i+l} - P_i)$. On average $\langle \epsilon_i (P_{l+i} - P_i) \rangle|_{V_i} \equiv \mathcal{R}(l, V_i)$. If ϵ_i is the sign of the market order, the transaction price is $P_i + \epsilon_i \frac{S}{2}$, where P_i is the mid-point price just before

[11] Assuming that he participates alone to a fraction γ of the trades number lead to the same conclusions.

the market order V_i is emitted. The gain $\mathcal{G}$ of the market maker can thus be written:

$$\mathcal{G} = \gamma \sum_{i=1}^{T} \epsilon_i V_i (P_i + \epsilon_i \frac{S_i}{2}) - \gamma (\sum_{i=1}^{T} \epsilon_i V_i) P_T \tag{3.57}$$

$$= \gamma \frac{1}{2} \sum_{i=1}^{T} V_i S_i - \gamma \sum_{i=1}^{T} \epsilon_i V_i (P_T - P_i) \tag{3.58}$$

Using that $\langle \epsilon_i (P_{\ell+i} - P_i) \rangle|_{V_i} \equiv \mathcal{R}(\ell, V_i)$, one obtains after averaging:

$$\frac{\langle \mathcal{G} \rangle}{\gamma} = \frac{1}{2} \langle VS \rangle_V T - \sum_{i=1}^{T} V_i \mathcal{R}(T-i, V_i) \tag{3.59}$$

We introduce the notation $\sum_{i=1}^{T} \langle \mathcal{R}(T-i, V_i) \rangle_V \equiv T \mathcal{R}_T$. If the average impact function $\mathcal{R}(\ell, V)$ did not vary with the trading time ℓ, we would obtain that $\langle V \mathcal{R}_T(V) \rangle_V = \langle V \mathcal{R}(V) \rangle_V$. Then Eq.(3.59) becomes:

$$\frac{\langle \mathcal{G} \rangle}{\gamma} = \frac{1}{2} \langle VS \rangle_V T - T \langle V \mathcal{R}(V) \rangle_V \tag{3.60}$$

Imposing that the gain is positive leads to Eq.(3.56).

As we showed empirically $\langle \mathcal{R}(\ell, V) \rangle_V \equiv \mathcal{R}(\ell)$ is in fact not constant. It varies non-monotonically with time: it starts at a finite value $\mathcal{R}(\ell = 1)$, grows roughly by a factor 2, and then decays again, see for example Figs.3.12 and 3.13. Thus the market maker average gains a priori depend on his horizon T, and Eq.(3.56) must be written using $\mathcal{R}_T$ instead of $\mathcal{R}$. In what follows we aim to test Eq.(3.56) using empirical data. In order to do this, we define $\lambda(T) = \mathcal{R}_T / \mathcal{R}(\ell = 1)$. Since empirically $\mathcal{R}(1) < \mathcal{R}(\ell) < 2\mathcal{R}(1)$ for all ℓ, one has $1 < \lambda(T) < 2$. Thus we predict that:

$$\langle SV \rangle_V \geq 2\lambda(T) \langle V \mathcal{R}(\ell = 1, V) \rangle_V \quad \text{with } 1 < \lambda(T) < 2 \tag{3.61}$$

To test Eq.(3.61), we average $\langle SV \rangle_V / \langle V \rangle_V$ and $\langle \mathcal{R}(\ell = 1, V) V \rangle_V / \langle V \rangle_V$ every 10000 trades (2 days) for France Telecom 2002. Doing so we obtain quantities which vary by a factor 5. This allows us to test the linear dependence of Eq.(3.61). Our result shown in Fig.3.23 is in good agreement

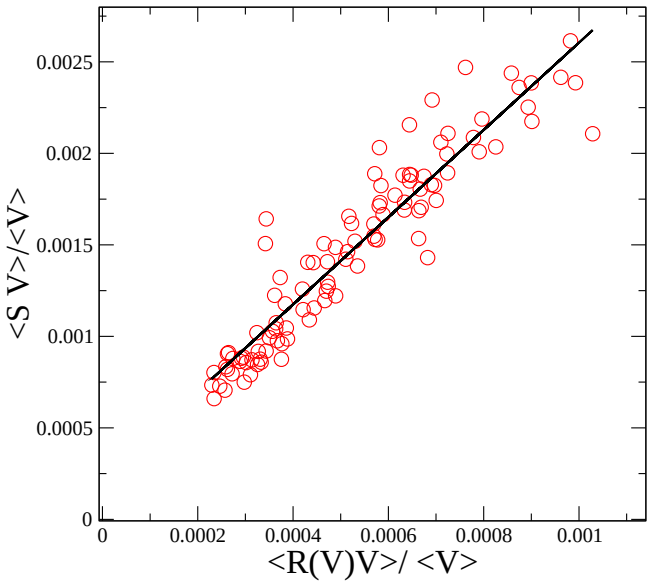

Fig. 3.23: $\langle SV\rangle_V/\langle V\rangle_V$ *vs* $\langle \mathcal{R}V\rangle_V/\langle V\rangle_V$ for France Telecom in 2002. Each dot corresponds to a pair $(\langle SV\rangle_V/\langle V\rangle_V,\ \langle \mathcal{R}V\rangle_V/\langle V\rangle_V)$ computed by averaging on 10000 trades. S and $\mathcal{R}$ are in relative value. The line is a fit and leads to $y = 0.00022 + 2.38x$. The tick size is, in this unit, 0.0006.

with the theoretical prediction. We find a clear linear dependence, and a slope 2.38 corresponding to $\gamma = 1.19$. The fit has a finite value at the origin, as expected since the spread cannot be smaller than the tick size, equal to 0.0006 in this unit.

We also test Eq.(3.61) using 60 different stocks of Paris Bourse with different capitalization. The relative values of the spread and the average impact can vary by a factor 5 between the different stocks, which enables to test the linear relation (3.61). Once again we find a good agreement with the predicted linear dependence, as shown in Fig.3.24. The fit leads to $y = 3.08x - 0.0002$. This corresponds to $\gamma = 1.5$.

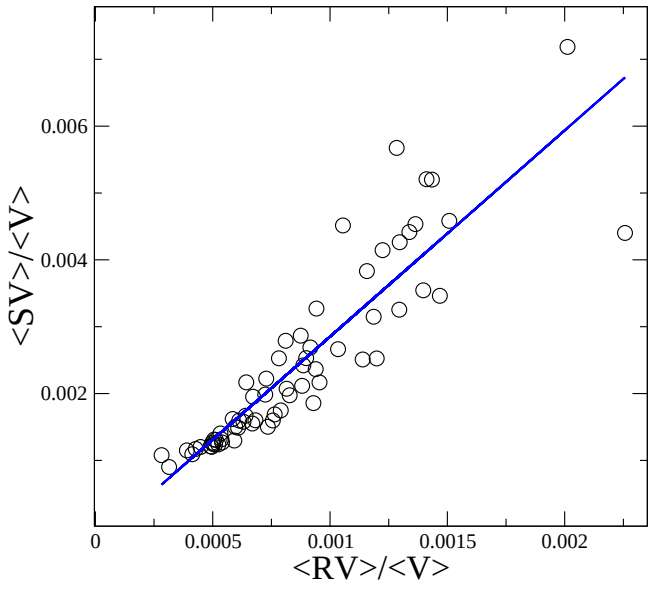

Fig. 3.24: $\langle SV\rangle_V/\langle V\rangle_V$ *vs* $\langle \mathcal{R}V\rangle_V/\langle V\rangle_V$ for 60 stocks of Paris Bourse in 2002. Each dot corresponds to a pair $(\langle SV\rangle_V/\langle V\rangle_V, \langle \mathcal{R}V\rangle_V/\langle V\rangle_V)$ computed for one stock over the year by averaging. S and $\mathcal{R}$ are in relative value. The line is a fit and leads to $y = 3.08x - 0.0002$.

A simplified version of Eq.(3.56) can be recovered with a "local" argument which does not involve any time-horizon. We consider another market making strategy where limit orders are deposited alternatively. At the begin-

ning there are deposited at the ask only, until a market order to buy comes. Then limit orders are deposited only at the bid. More generally, each time limit orders from the market maker are traded, limit orders are put only on the other side of the order-book. With this strategy there are no inventory problems. Assuming for simplicity that all volumes have the same size, and that there are no sign correlations in the market orders, it is simple to compute the gain of this strategy. There is a probability 1/2 that a sell market order follows a buy market order $(\epsilon_1, \epsilon_2) = (+, -)$, in which case the gain is $S - \mathcal{R}$ (in this simple model $\mathcal{R}$ does not display time dependence). With a probability 1/4, we have (+,+,-) and the gain is $S - 2\mathcal{R}$. Generalizing we have:

$$\langle \mathcal{G} \rangle = \sum_{n \geq 1} (S - n\mathcal{R}) \{\frac{1}{2}\}^n = S - 2\mathcal{R} \tag{3.62}$$

which corresponds to Eq.(3.56) when fluctuations of volume are absent. In economical terms, the equality corresponding to Eq.(3.62) has a very simple meaning: it indicates that on average, the new price $P_F = P_i + \mathcal{R}$ is the last transaction price $P_i + S/2$.

3.4.2 Liquidity vs. volatility

In section 3.3, by assuming that the price dynamics was diffusive —as it is the case—, we obtained the relation (3.48) between the diffusion constant and the price impact function, that we may write:

$$\frac{\mathcal{D}(\ell)}{\ell} = A_0 \langle \mathcal{R}^2(V) \rangle_V + D_\eta \tag{3.63}$$

where D_η corresponds to the part of the price diffusion that does not involve market orders, and A_0 is a constant that depends on the temporal correlation of the market order flow. In what follows we aim to use Eq.(3.63) and Eq.(3.56) to show that volatility per trade and spread are proportional, and that this linear dependence can be observed by considering stocks of different capitalization, or by considering the time evolution of the spread of a single stock. In order to do so, we must also discuss how A_0 and D_η can depend on capitalization or time.

In what follows we shall assume that the diffusion induced by limit orders D_η contributes to a fixed fraction of the total diffusion constant: $D_\eta = A_3 \langle \mathcal{R}^2(V) \rangle_V$, where A_3 is a constant. For FT, D_η represents of the order of 1/2 of the total price diffusion, as it appears to be from Fig.(3.20). We assume that it is also the case for other stocks.

A possible justification for this is the following: our empirical results support that limit and market orders have similar costs. This suggests that

both are used equivalently by agents, and that limit orders and market orders impact prices in a similar way and are responsible each for roughly half of the total volatility. Note that this assumption can be relaxed: if the fraction of the total volatility that D_η represents differs between stocks, but does not depend systematically of the spread nor the volatility per tick, then the term D_η would simply add noise to our results below.

Concerning A_0, we shall assume for simplicity that it does not vary between stocks. Our results hold as long as this quantity does not display systematical dependence with the spread. In the framework of a perfectly diffusive price discussed at the end of the section 3.3.3, one finds that $A_0 \sim K(0)^2$. $K(0)$ depends on the market order sign correlation $C_0(\ell)$, and mostly of the difference $C_0(0) - C_0(1)$. In what follows we shall assume that this quantity does not vary systematically with the capitalization.

For the volatility per trade $\sigma_1 \equiv \sqrt{\mathcal{D}(1)}$ we have, following these approximations and introducing $B = \sqrt{A_0 + A_3}$:

$$\sigma_1 = B\langle \mathcal{R}^2(V)\rangle^{1/2} \tag{3.64}$$

As we discussed in the last paragraph, in a competitive market (if ticks are not too large) another constraint appears, which characterizes the spread: $\langle VS\rangle_V = 2\langle \mathcal{R}(V)V\rangle_V$. It turns out to be convenient to use this expression to write:

$$\sigma_1 = S\{2B\frac{\langle \mathcal{R}^2(V)\rangle_V^{1/2}\langle VS\rangle_V}{\langle \mathcal{R}(V)V\rangle_V S}\} \tag{3.65}$$

This equation relates liquidity (S) and volatility. We argue below that the term in bracket behaves as a constant for stocks with different capitalization, or when the order-book evolves in time. In a simple model where all market orders have the same size, the distribution of volume is a delta-function and we find obviously the desired linear dependence between S and σ:

$$\sigma_1 = 2BS \tag{3.66}$$

Note that the same assumption applied to Eq.(3.56) leads to $S = 2\mathcal{R}$. Fig.3.25 shows that this assumption gives the right linear dependence, but not the right slope. In what follows we aim to show that this statement is also true for Eq.(3.65), and that the term into bracket is indeed constant, and does not vary with capitalization nor time.

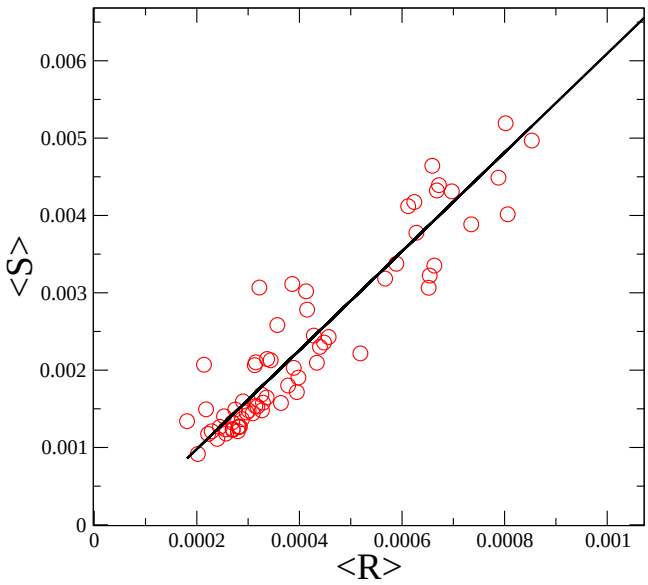

Fig. 3.25: $\langle S \rangle$ *vs* $\langle \mathcal{R} \rangle$ for France Telecom in 2002. Each dot corresponds to a pair $(\langle S \rangle, \langle \mathcal{R} \rangle)$ computed by averaging on 10000 successive trades. $\langle S \rangle$ and $\langle \mathcal{R} \rangle$ are in relative value. The line is a linear fit and leads to $y = -0.0003 + 6.4 * x$.

In order to show this, we make the following hypothesis of "universality": if one uses the proper re-scaling (or change of units) of the variables that characterize different continuous double-auctions, these systems display identical properties. As we discussed in the introduction, this universality was supported empirically for the price impact and the average order-book depth. In practice, we shall assume here that for all quantities that depends on the market order volume V, such as the price impact or the size distribution of market orders, the dependence with V only appears as $V/\bar{V}$, where $\bar{V}$ is the average market order size, which contains the dependence on capitalization (Zumbach, 2004) and time.

Specifically, we assume that the distribution of market order size $P(V)$ can be written as :

$$P(V) = \frac{1}{\bar{V}} f(V/\bar{V}) \tag{3.67}$$

where $\bar{V}$ is the characteristic size of the market orders, and f is a function which does not depend on the stock considered, and is called a "scaling function". For the price impact we write:

$$\mathcal{R}(V) = \frac{1}{L} g(V/\bar{V}) \tag{3.68}$$

where g is a scaling function, and $L = L(C,t)$ characterizes the liquidity of a stock, and can depend a priori of capitalization and time. Finally we assume that the probability $P(S;V)$ of finding S and V writes:

$$P(V;S) = h(S/\bar{S}; V/\bar{V}) \tag{3.69}$$

where $\bar{S}$ is the average spread, which can vary with capitalization or time. It is straightforward to show that this equations leads to $\langle SV \rangle = A_1 \bar{S}\bar{V}$ where A_1 is a numerical factor. Computing Eq.(3.65) by introducing $u = V/\bar{V}$, one obtains:

$$\frac{\sigma_1}{\bar{S}} = 2BA_1 \frac{(\int_0^\infty f(u)g^2(u)du)^{1/2}}{\int_0^\infty f(u)ug(u)du} \equiv A_2 \tag{3.70}$$

which shows that the ratio $\sigma_1/\bar{S}$ does not depend on $\lambda(C)$ nor on $\bar{V}$, and therefore does not depend on the market capitalization. Thus the average spread $\bar{S}$ is therefore directly proportional to the volatility per trade:

$$\sigma_1 = A_2 \bar{S} \tag{3.71}$$

as observed (see Fig 1.1). The two constraints that the price must be diffusive, and that the gain of limit and market orders are equal, lead to this relation between liquidity and volatility. This result shed lights on long-standing questions of the microstructure literature. It shows that if both

the trading activity and the volatility *per unit time* increase, the spread can remain the same if the volatility *per trade* is constant. Conversely, if the trading frequency ν increases, while the volatility per unit time is constant, then the spread decreases as $S \sim \nu^{-1/2}$, and the liquidity improves. This situation occurs for example with the FTSE 100: the spread diminishes with the different stock capitalization C because (i) the volatility per unit time does not vary much with C and (ii) the trading frequency ν increases with capitalization as $\nu \sim C^{0.44}$ (Zumbach, 2004). According to our result this leads to $S \sim C^{-0.22}$, in good agreement with the empirical data.

3.5 Conclusion and Perspectives

3.5.1 Summary

In this work we focused on three distinct points of market microstructure.

We studied theoretically some "zero intelligence" models where agents behavior are mimicked by random flows of order. In particular we furnished qualitative arguments to rationalize the behavior of price diffusion and spread, that were previously computed numerically (Farmer, Patelli, et al., 2004). We proposed a new, more realistic model. An interesting question that comes out of this approach is the very definition of price. We showed empirically that it is indeed possible to build a proxy of the price more natural than the mid-point, which takes into account the instantaneous repartition of offer and demand. Concerning other market properties, we ultimately argued that the zero-intelligence approach is not appropriate to describe price formation. It completely misses the fact that forbidding simple statistical arbitrage strategies constrains much the order-book properties.

Our second point was to show that the market displays *long memory*: the correlation of the market order signs decay in time as a non-integrable power-law, which lasts *days*. This implies that the price impact of one *single* trade is not permanent, but rather almost completely transient, and decays slowly in time. This is a necessary condition for the price to be diffusive. We argued that this decay is induced by the emission of "contrarian" limit orders. We showed how a simple liquidity providers strategy can lead to a purely diffusive price. Such equilibrium between market orders and limit orders is reached if liquidity providers change price proportionally to the *innovation*, i.e. the *unpredictable* content of the last market order.

Finally, we studied how liquidity and volatility are related. Our main result is to show that, without assuming any underlying model for the risk aversion of agents, nor for the information content of trades , it is possible to

relate analytically spread and average price impact. Our single assumption is to impose that market orders and limit orders have on average an equal gain, as it should be the case in a continuous double-auctions set-up where both types of order can be emitted by any agent. We checked empirically this relation, and found a good agreement with our predictions. Since volatility and market impact are related, we obtained that volatility per tick and spread are proportional. We furnished new data on this relation. Our work enables to understand how liquidity, volatility and trading activity are related, a much debated question in the microstructure literature.

3.5.2 Perspectives

Other scaling properties?

It is tempting to try to derive other relations among micro-structural properties. To discuss this possibility, we introduce the following parameters: ν is the frequency of market orders, σ is the volatility per unit time, $\sigma_1 = \sigma\nu^{-1/2}$ is the volatility per trade, V_l is the volume of limit orders in the order-book, V_m is the volume of market orders, S is the spread and $\mathcal{R}$ is the price impact. Previously we established that:

$$\sigma \equiv \sigma_1\nu^{1/2} \sim \mathcal{R}^{1/2}\nu^{1/2} \tag{3.72}$$
$$S \sim \mathcal{R} \tag{3.73}$$
$$\sigma \sim S\nu^{1/2} \tag{3.74}$$

Recently, Zumbach (2004) studied empirically how the different quantities we introduced above scale with the company size, or capitalization C in the FTSE 100. He finds, by using power law fits:

$$\nu \sim C^{0.39} \tag{3.75}$$
$$V_m \sim C^{0.44} \tag{3.76}$$
$$V_l \sim C^{0.76} \tag{3.77}$$
$$\sigma \sim C^{0} \tag{3.78}$$
$$S \sim C^{-0.23} \tag{3.79}$$
$$V_m\nu \sim C^{0.9} \tag{3.80}$$
$$S \sim \sigma_1^{0.94} \tag{3.81}$$

Similar exponents (although with some differences, in particular for the dependence of ν) can be observed with Paris Bourse stocks (Kockelkoren et al., 2005). Eq.(3.78) indicates that small and large companies have similar volatility. It is a surprising result since the fluctuation of the company growth

decay with capitalization (Wyart & Bouchaud, 2003). Eq.(3.80) indicates that the volume traded is roughly proportional to the capitalization. Why it should be so is not entirely obvious. A possible justification is that a fraction of the agents manage portfolios, and track the index. Doing so they trade proportionally to the capitalization of each stock. If Eq.(3.80) is accepted, Eq.(3.75) leads to (3.76). Finally, Eq.(3.79) and (3.81) are good approximations of our results (3.72), once (3.75) and (3.78) are accepted.

Obviously a useful result to understand how price forms, and in particular what fixes the absolute value of the spread, would be to derive Eqs.(3.75) and (3.77). Nevertheless these relations are probably not strict constraints stemming from the absence of simple statistical arbitrage, such as those of Eqs.(3.72): if it were so, the order-book would be completely constrained, and would not evolve with time. Some of these relations might also be related to market features exogenous of the microstructure, such as the dependence with C of the average repartition of a company between its share holders. Finally, they could be pure conventions, in the sense of Chapter II.

Order-book dynamics

Our description above is purely static. Another curious feature of the continuous double-auctions order-books is the fact that they fluctuate a lot with time. In what follows we document the "seasonal" evolution of the order-book: every day, the properties of price formation, such as the parameters discussed in the last paragraph, display regular patterns. For example, the volatility and the traded volume of the Paris Bourse stocks are high in the morning and the evening, and much smaller at noon. Such patterns exist, and differ, in all markets. There existence is interesting for the following reason: they furnish a way to study which constraints characterize order-books. For example, Eqs.(3.72) must be satisfied throughout the day: if not, statistical arbitrage would be possible. Thus quantities that evolve differently during the day are not bound to each other.

Fig.(3.26) shows the daily patterns of FT. We observe that: (i) the spread and volatility per tick display very similar curves, as expected. These quantities decrease by a factor 2 between 9 and 10 a.m, and then stay nearly constant throughout the day. (ii) The average transaction volume V_m is constant, whereas the volume in the order-book increases of a factor 2. This implies that, somewhat surprisingly, these 2 quantities are not bound to each other. This also shows that, curiously enough, the order-book depth can increase, while the market order size and the spread, and therefore the price impact, stay constant. (iii) The volatility per unit time σ strongly changes, and is roughly 4 times larger at the market closing and opening than at

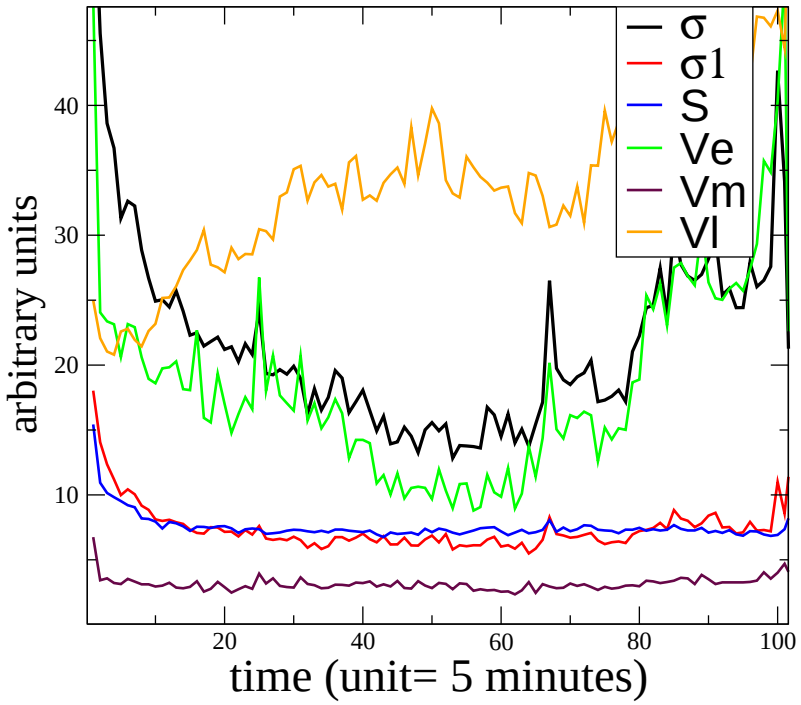

Fig. 3.26: Daily patterns of the volatility σ, the volatility per trade σ_1, the spread S, the total volume exchanged V_e, the average transaction volume V_m, and the characteristic volume in the order-book $V_l = V_a + V_b$ (V_a and V_b are the volume at ask and bid) *vs* time for France Telecom. These quantities are computed every 5 minutes, and the curves obtained are averaged over one year. The day starts at 9 a.m. and ends at 5 p.m, which leads to 102 points.

noon. (iv) the volume exchanged per unit time V_l is qualitatively similar to the volatility: it decreases by a factor 4 in the morning and increases again in the afternoon.

This results indicate that the seasonal fluctuations of volatility are mostly induced by fluctuations in the trading frequency, since the volatility per trade stays almost constant (except in the morning). Thus most of the seasonal volatility fluctuations could be flatten by a proper rescaling of time. Nevertheless, such a rescaling does not explain (i) the morning change of volatility per trade or spread and (ii) the growth of the order-book depth during the day. These two effects correspond to distinct distortions of the order-book.

To conclude, the quantitative understanding of the order-book is in its infancy. In particular it is yet unclear if other constraints, such as those of between spread, volatility, and response, exist between the other quantities that characterize price formation. The observed distortions of the order-book rather suggest that it has several unfixed degrees of freedom. It would be interesting to determine the class of order-book distortions which are allowed without generating statistical arbitrage opportunities.

3.6 References

Ang A., Bekaert G., 2002. International Asset Allocation with Regime Shifts. The Review of Financial Studies 15, 1137-1187.

Arthur W. B., 1995. Complexity in Economic and Financial Markets. Complexity 1, 20-25.

Bak P., 1996. How Nature Works: The Science of Self-Organized Criticality. Copernicus, Springer, New York.

Banerjee A., 1992. A simple model of herd behavior. Quarterly Journal of Economics 107, 797-817.

Barberis N., Shleifer A., Vishny R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307-343.

Barra, 1997, Market Impact, unpublished report.

Beja A., Goldman M. B., 1980. The dynamics of prices in disequilibrium. Journal of Finance 35, 235-248.

Biais B., Foucault T. and Hillion P, 1997 Microstructure des marchés financiers, PUF

Bikhchandani S., Hirshleifer D., Welch I., 1992. A theory of fads, fashions, custom and cultural changes as informational cascades. Journal of Political Economy, 100 992-1026

Black F., 1986. Noise. Journal of Finance, 41 529-543.

Black A., Fraser P., 2000. Are stock prices too volatile and returns too high? A reassessment of the empirical evidence using a dynamic version of the CAPM, Working Paper, Aberdeen.

Bouchaud J.-P., Cont R., 1998. A Langevin approach to stock market fluctuations and crashes. European Journal of Physics B 6, 543-550.

Bouchaud J.-P., Hasareesing A., 2000. Barrier crossing with non Gaussian tailed noise. Mimeo, unpublished.

Bouchaud J.-P., Giardina I., Mézard M., 2001. On a universal mechanism for long ranged volatility correlations. Quantitative Finance 1, 212-216.

Bouchaud J.-P., Potters M., 2003. Theory of Financial Risks and Derivative Pricing. Cambridge University Press, 2000.

Bouchaud J.-P., Gefen Y., Potters M., Wyart M., 2004. Fluctuations and response in financial markets: The subtle nature of 'random' price changes. Quantitative Finance 4, 176-190

Bouchaud J.-P., Kockelkoren J., Potters M., 2004. Random walks, liquidity molasses and critical response in financial markets. Working paper, available at xxx.lanl.gov/pdf/cond-mat/0406224.

Brière M, 2005. Formation des Taux d'Interet, Anomalies et Croyances Collectives, Economica, Paris

Brock W., Hommes C., 1997. A rational route to randomness, Econometrica 65, 1059-1096.

Challet D., Chessa A., Marsili M., Zhang Y.-C., 2001. From Minority Games to real markets. Quantitative Finance, 1, 168-176.

Challet D., Marsili M., 2003. Criticality and Finite size effects in a simple realistic model of stock markets. Physical Review E 68, 036132-36.

Challet D., Marsili M., Zhang Y.-C., 2004. The Minority Game. Oxford University Press.

Chandrasekhar S., 1943. Stochastic problems in Physics and Astronomy. Review of Modern Physics 15, 1-89.

Chordia, Tarun, Avanidhar Subrahmanyam, and Ravi Anshuman, 2001, Trading Activity and Expected Stock Returns, Journal of Financial Economics **59**, 3-32

Chiarella C., He X.-Z., 2001. Asset price and wealth dynamics under heterogeneous expectations. Quantitative Finance 1, 509-526.

Cochrane J. H., Campbell J., 1991. Volatility tests and efficient markets: A review essay. Journal of Monetary Economics 27 463-485.

Cochrane J. H., Campbell J., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior, Journal of Political Economy 107, 205-251 (1999).

Cont R., Bouchaud J.P., 2000. Herd behavior and aggregate fluctuations in financial markets. Macroeconomic Dynamics 4, 170-195.

Cont R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1, 223-236.

Daniels M., Farmer J.D., Iori G., Smith E., 2001, Demand storage, market liquidity, and price volatilty, SFI working paper 02-01-001. This paper appeared in final, but truncated form as: M. G. Daniels, J. D. Farmer, G. Iori, E. Smith, Quantitative model of price diffusion and market friction based on trading as a mechanistic random process, Phys. Rev. Lett. **90**, 108102 (2003).

De Bondt W., Thaler R., 1985. Does the market overreact ? Journal of Finance 40, 793-805.

Ding Z., Granger C. W. J., Engle R. F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

Fama E. F., 1970. Efficient capital markets: A review of theory and empirical work. Journal of Finance 25, 383-417.

Farmer J. D., 2002. Market Force, Ecology and Evolution. Industrial and Corporate Change 11, 895-953.

Farmer D., Gillemot L., Lillo F., Mike S., and Sen A., 2004, What really causes large price changes? Quantitative finance, vol. 4, 383-397

Farmer J.D., Patelli P., and Zokvo I., 2004, The predictive power of zero intelligence in Financial Markets

Fisher A., Calvet L., Mandelbrot B. B., 1997. Multifractality of DEM/$ rates. Cowles Foundation Discussion Paper 1165

Föllmer H., Horst U., Kirman A., 2005. Equilibria in financial markets with heterogeneous agents, a new perspective. Journal of Mathematical Economics 41 123-155.

Gabaix X., Gopikrishnan P., Plerou V. and Stanley G.H., 2003, A theory of power laws distributions in financial markets fluctuations, Nature **423**, 267

Gardiner C. W., 1996. Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences (2nd edition). Springer Series in Synergetics, Vol 13, Springer Verlag.

Ghashghaie S., Breymann W., Peinke J., Talkner P., Dodge Y., 1996. Turbulent cascades in foreign exchange markets. Nature, 381, 767-769.

Giardina I., Bouchaud J.-P., 2003. Bubbles, Crashes and Intermittency in agent based market models. European Journal of Physics B 31, 421-437.

Goldenfeld N., 1992. Lectures on phase transitions and critical phenomena. Frontiers in Physics, Westview Press.

Guedj O., Bouchaud J.-P., 2005. Experts earning forecasts: bias, herding and gossamer information. International Journal of Theoretical and Applied Finance, in the press.

Guillaume D. M., Dacorogna M. M., Davé R. D., Müller U. A., Olsen R. B., Pictet O. V., 1997. From the bird's eye to the microscope. Finance and Stochastics 1, 95-129.

Hasbrouck J., 1991,Measuring the information content of stock trades, Journal of Finance, **XLVI**, 179

Hommes C., 2001. Financial markets as nonlinear adaptive evolutionary systems. Quantitative Finance 1, 149-167.

Hong H., Stein J. 1999. A unified theory of underreaction, momentum trading and overreaction in asset markets. Journal of Financial Economics. 54, 2143-2184.

Hopman, C. 2002 *Are supply and demand driving stock prices?*, MIT working paper.

Iori G., 1999. Avalanche Dynamics and Trading Friction Effects on Stock Market Returns. International Journal of Modern Physics C 10, 1149-1162.

Jefferies P., Hart M., Hui P. M., Johnson N. F., 2000. Trade dynamics in a model market. International Journal of Theoretical and Applied Finance 3, 443-450.

Kahneman D., Tversky A. 1986. Choices, values, and frames. American Psychologist, 341-350.

Kearns M. and Ortiz L., 2003, The Penn-Lehman Automated Trading Project, Agents and Markets, 22-31

Kempf A., Korn O., 1999. Market Depth and Order Size, Journal of Financial Markets 2, 29-48.

Keynes J. M., 1936. The general theory of Employment, Interest and Money. McMillan, London (1936).

Kirman A., 1993. Ants, rationality and recruitment. Quarterly Journal of Economics 108, 137-156.

Kirman A., Teyssière G., 2002. Microeconomic models for long range memory in the volatility of financial time series. Studies in Non-Linear Dynamics and Econometrics 5, 281-302.

Kockelkoren et al., 2005, in preparation

Kyle A. S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1335.

LeBaron B., 2001. Stochastic volatility as a simple generator of apparent financial power laws and long memory. Quantitative Finance 1, 621-631.

LeBaron B., 2002. Calibrating an agent based financial market to macroeconomic time series. Technical report, Brandeis University (2002).

Lenzi E. K., Anteneodo C., Borland L., 2001. Escape time in anomalous diffusive media. Physical Review E 63, 051109-13.

Lillo F., Farmer J. D., Mantegna R., 2003. Master curve for price-impact function. Nature 421, 129-130.

Lillo F., Farmer J.D., 2003. On the origin of power-law tails in price fluctuations, e-print cond-mat/0309416

Lillo F., Farmer J.D., 2003. The long memory of efficient markets, Preprint e-print cond-mat/0311053

Liu Y., Cizeau P., Meyer M., Peng C.-K., Stanley H.E., 1997. Correlations in Economic Time Series. Physica A245, 437-440

Lo A., 1991. Long term memory in stock market prices. Econometrica 59, 1279-1313.

Longin F., Solnik B., 2001. Extreme Correlation of International Equity Markets. Journal of Finance 56, 651-678.

Lux T., Marchesi M., 1999. Scaling and criticality in a stochastic multi-agent model, Nature 397, 498-500 (1999).

Lux T., Marchesi M., 2000. Volatility Clustering in Financial Markets: A Microsimulation of Interacting Agents. International Journal of Theoretical and Applied Finance 3, 675-702.

Lux T., 2001. Turbulence in financial markets: the surprising explanatory power of simple cascade models. Quantitative Finance, 1, 632-640.

Mandelbrot B. B., 1997. Fractals and Scaling in Finance. Springer, New York.

Maslov S. (2000), Simple model of a order-driven market, Physica A, **278**, 571

Mendelson H., 1982, Market behavior in a clearing house, Econometrica **50**, 1505-1524 (1982)

Muzy J.-F., Delour J., Bacry E., 2000. Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. European Journal of Physics B 17, 537-548.

Nagel R, 1995. Unraveling in Guessing Games: An experimental study. American Economic Review, 85 1313-1326.

Orléan A., 1999. Le pouvoir de la finance. Odile Jacob, Paris.

Palmer R. G., Arthur W. B., Holland J. H., LeBaron B., Taylor P. 1994. Artificial economic life: a simple model of a stock market. Physica D75, 264-274.

Pastor, L. and Stambaugh, F., 2003, Liquidity Risk and Expected Stock Returns,? Journal of Political Economy, **111** , 642-685.

Percival D.B., 1992, Simulating Gaussian random processes with specified spectra, Comp. Sci. Stat. **24** 534

Plerou V., Gopikrishnan P., Gabaix X., Stanley H. E., 2002. Quantifying Stock Price Response to Demand Fluctuations. Physical Review E 66, 0271041-44.

Potters M., Cont R., Bouchaud J.-P., Financial markets as adaptive ecosystems. Europhysics Letters 41, 239-244.

Potters M., Bouchaud J.-P., 2003. More statistical properties of order books and price impact. Physica A324, 133-140.

Rosenow B., 2002, Fluctuations and market friction in financial trading, Int. J. Mod. Phys. C, **13**, 419

Samuelson P. A., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial Management Review 6, 41-49.

Sherstov A., Stone P., Three Automated Stock-Trading Agents: A comparative Study

Shiller R. J., 1981. Do Stock Prices move too much to be justified by subsequent changes in dividends ? American Economic Review 71, 421-436.

Shiller R. J., 2000. Irrational Exuberance. Princeton University Press.

Shiller R. J., 2002. From Efficient Market Theory to Behavioral Finance. Cowles Foundation working paper # 1385

Shleifer A., 2000. Inefficient Markets, An Introduction to Behavioral Finance, Oxford University Press.

Smith E., Farmer D., Gillemot L. and Krishnamurthy S, 2003, Statistical theory of the continuous double auction, Quant. finance, **3** 481-514

Trueman B., 1994. Analyst Forecasts and Herding Behavior. Review of Financial Studies, 7, 97-124.

Woodford M. 1990. Learning to believe in sunspots. Econometrica, 58, 277-307.

Weber P., Rosenow B., 2003, Order book approach to price impact, eprint: cond-mat 0311457

White Marc, 2004 A research Agenda for Agent-based Finance Models, working paper, www.geocities.com/wallstreet/7891/agenda.htm?200412

Wyart M, Bouchaud J-P, Self-referential behavior, overeaction and conventions in financial markets, accepted in Journal of Economical and Behavioral Organization

Wyart M, Bouchaud J-P, 2003, Statistical Model for company growth, Physica A **326**, 241

PHYSA 7522
pp: 1-15 (col.fig.: Nil)
PROD. TYPE: COM
ED: Rohini
PAGN: VD – SCAN: Ravi

ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

PHYSICA A

ELSEVIER

Physica A ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

www.elsevier.com/locate/physa

1

Statistical models for company growth

3 Matthieu Wyart[a], Jean-Philippe Bouchaud[a,b,*]

[a] *Service de Physique de l'État Condensé, Orme des Merisiers—CEA Saclay,*
5 *91191 Gif sur Yvette Cedex, France*
[b] *Science and Finance, Capital Fund Management 109-111 rue Victor-Hugo,*
7 *92532 Levallois Cedex, France*

Received 19 November 2002; received in revised form 26 February 2003

9 **Abstract**

We study Sutton's 'microcanonical' model for the internal organization of firms, that leads to
11 non-trivial scaling properties for the statistics of growth rates. We show that the growth rates are asymptotically Gaussian in this model, whereas empirical results suggest that the kurtosis of the
13 distribution increases with size. We also obtain the conditional distribution of the number and size of sub-sectors in Sutton's model. We formulate and solve an alternative model, based on the
15 assumption that the sector sizes follow a power-law distribution. We find in this new model both anomalous scaling of the variance of growth rates and non-Gaussian asymptotic distributions.
17 We give some testable predictions of the two models that would differentiate them further. We also discuss why the growth rate statistics at the country level and at the company level should
19 be identical.

21 *PACS:* ■; ■; ■

Keywords: ■; ■; ■

23 1. Introduction

The annual growth rate of a company is fluctuating both across companies and from
25 year to year. It is, therefore, tempting to study the statistics of this growth rate. It has
27 been known for many years that the *average* growth rate is, to a good approximation,

* Corresponding author. Service de Physique de l'Etat Condense, CEA-Saclay, Orme des Merisiers, Gif sur Yvette Cedex 91191, France.
E-mail address: bouchaud@spec.saclay.cea.fr (J.-P. Bouchaud).

0378-4371/03/$ - see front matter © 2003 Published by Elsevier Science B.V.
doi:10.1016/S0378-4371(03)00267-X

independent of the size of the company. This is known as Gibrat's proportionality law: since the growth rate is the *relative* size increase of a company (where the size refers to the sales, the number of employees, etc.), the fact that the average growth rate is independent of the size means that on average a company grows proportionally to its size.

A very interesting question, that was only addressed recently, concerns the fluctuations of the growth rate, and the size dependence of these fluctuations. Quite remarkably, Stanley et al. found that the standard deviation σ of the growth rate r decreases with the size S of the company as $\sigma(S) \sim S^{-\beta}$, with $\beta \approx 0.15$ [1,2]. This power-law scaling holds over six decades, and can be extended to larger sizes by considering countries as 'companies' and taking the GNP as a measure of the size [3]. More precisely, the distribution of the rescaled growth rate $v = r/\sigma(S)$, with $\sigma(S) \sim S^{-\beta}$, appears to be size independent. This rescaled distribution $\Pi(v)$ is furthermore found to be non-Gaussian. A further investigation of the empirical evidence by Sutton gives a value of β that fluctuates in the interval $[0.15, 0.21]$ [4].

This remarkable result is puzzling because one could have naively expected that large companies (or countries for that matter) would aggregate different independent 'shocks' that would lead, using the central limit theorem, to a $S^{-1/2}$ decrease of the volatility of its growth rate, which would furthermore be Gaussian for large S's. This however assumes that a company can be thought of as a collection of K 'sub-companies' of average size S_0 and weakly correlated activities. In this case, $K = S/S_0$ and if the shocks affecting each sector of activity have a finite second moment, the central limit theorem applies.

The fact that $\beta < \frac{1}{2}$ suggests otherwise. Obviously, if all the sectors of activity of a given company had strong cross-correlations, one would find the extreme result that $\beta = 0$. However, this is not the case: Sutton has shown some empirical data that support the idea that the growths of different sectors are to a good approximation uncorrelated [4]. This is what Sutton called the 'scaling puzzle', which lead him to propose a simple model for the internal organization of firms that predicts asymptotically $\beta = \frac{1}{4}$, not very far from the empirical result [4]. Actually, Sutton shows that for finite S, his model predicts an effective value of β is slightly below $\frac{1}{4}$.

The aim of this note is threefold. In the first part which we intend to be also of pedagogical interest, we revisit Sutton's model using methods from statistical physics, and obtain a number of complementary predictions of this model that can be compared with empirical data, in particular the distribution of rescaled growth rate $\Pi(v)$, which we find to be asymptotically Gaussian. This must be contrasted with empirical results that suggest that the kurtosis of the distribution (that measures the distance from the Gaussian) *increases* with size. Second, we introduce and study an alternative model where we argue that the distribution of sizes of the sub-sectors is a power-law, and derive analytically the value of β and the shape of $\Pi(v)$, which in some regime is found to be strongly non-Gaussian. We then compare our results to the findings of Stanley et al. and discuss the plausibility of our alternative model. Finally, we discuss the interesting fact that GNP growth and company growth behave similarly. This means that the microeconomical and macroeconomical levels are strongly interconnected. We show that our model is indeed stable upon aggregation.

2. Sutton's model

We first recall Sutton's model. In the absence of more information, Sutton postulates that all partitions of a company of size S in smaller sub-pieces are equiprobable [4]. This is a kind of 'microcanonical', minimum information assumption, similar to the corresponding hypothesis in statistical physics where all microstates are equiprobable. For physical systems, this is justified by the Liouville theorem that is itself a consequence of Hamiltonian dynamics; it would be interesting to find an analogue of this theorem for the (stochastic) dynamics underlying the organization of firms.

More precisely, Sutton assumes that S is a large integer, and uses known mathematical results on the number of partitions to compute $\sigma(S)$. Let us show how his results can be recovered directly. For this, we introduce the following quantity: [1]

$$\mathcal{N}(R,K,S)=\sum_{s_1=1}^{\infty}\sum_{s_2=s_1}^{\infty}\cdots\sum_{s_K=s_{K-1}}^{\infty}\delta\left(S-\sum_{i=1}^{K}s_i\right)\int\prod_{i=1}^{K}P(\eta_i)\,\mathrm{d}\eta_i \times\delta\left(R-\sum_{i=1}^{K}s_i\eta_i\right)\,. \tag{1}$$

This quantity counts the number of partitions of the integer S in exactly K integers $s_1,s_2,\ldots,s_K$, and such that the total absolute growth rate R is given by the sum of independent random variables η_i (that we suppose of finite variance), each weighted by the size s_i of the sub-sector. This assumes a proportionality effect at the sub-sector level. It will be convenient to introduce the Fourier–Laplace transform of this quantity (or generating function), defined as

$$\hat{\mathcal{N}}(q,v,\lambda)=\sum_{S=1}^{\infty}\sum_{K=1}^{\infty}\int_{-\infty}^{\infty}\mathrm{d}R\,\exp[\mathrm{i}qR-vK-\lambda S]\mathcal{N}(R,K,S)\,. \tag{2}$$

The quantity $\hat{\mathcal{N}}(q=0,\ v=0,\ \lambda)$ is therefore the Laplace transform of the total number of partitions, and is given by

$$\hat{\mathcal{N}}(q=0,\ v=0,\lambda)=\sum_{K=1}^{\infty}\sum_{s_1=1}^{\infty}\sum_{s_2=s_1}^{\infty}\cdots\sum_{s_K=s_{K-1}}^{\infty}\exp\left(-\lambda\sum_{i=1}^{K}s_i\right) \tag{3}$$

and be computed explicitly as

$$\hat{\mathcal{N}}(q=0,\ v=0,\ \lambda)=\sum_{K=1}^{\infty}\frac{\mathrm{e}^{-\lambda K}}{\prod_{i=1}^{K}(1-\mathrm{e}^{-\mathrm{i}\lambda K})}\,. \tag{4}$$

For $\lambda\to 0$, the sum over K can be approximated by an integral

$$\hat{\mathcal{N}}(q=0,\ \mu=0,\ \lambda)\approx\int_0^{\infty}\mathrm{d}K\,\exp\left(-\lambda K-\int_0^{K}\mathrm{d}x\,\ln(1-\mathrm{e}^{-\lambda x})\right)\,. \tag{5}$$

[1] In the following equation, δ refers to the Dirac delta for continuous variables and to the Kronecker delta for discrete variables.

Now the integral over K can be estimated using Laplace's saddle-point method. The saddle point K^* obeys the following equation:

$$\lambda = -\ln(1 - e^{-\lambda K^*}) , \tag{6}$$

which for small λ gives

$$K^* \approx \frac{1}{\lambda} \ln \frac{1}{\lambda} . \tag{7}$$

Plugging this result in Eq. (5) leads to

$$\hat{\mathcal{N}}(q = 0,\ v = 0,\ \lambda \to 0) \sim \exp\left(\frac{1}{\lambda} \int_0^\infty dv \ln(1 - e^{-v})\right) = \exp\left(\frac{\pi^2}{6\lambda}\right) , \tag{8}$$

where we have neglected preexponential corrections, that can also be computed. Now, it is easy to check that the inverse Laplace transform of $\hat{\mathcal{N}}(q = 0,\ v = 0,\ \lambda \to 0)$ behaves, for large S, as

$$\mathcal{N}(S) \sim \exp[b\sqrt{S}], \quad b = \pi\sqrt{\tfrac{2}{3}} , \tag{9}$$

which is the Hardy–Ramanujan result at large S [5]. In the course of the calculation, one also discovers that, as far as scaling is concerned, $\lambda \sim S^{-1/2}$. One could extend the computation to get the exact prefactor, equal to $(4\sqrt{3}S)^{-1}$.

One can easily extend the computation to $v \neq 0$. The saddle point is now at

$$K^* \approx \frac{1}{\lambda} \ln \frac{1}{\lambda + v} . \tag{10}$$

Now, setting $v = x\lambda/|\ln \lambda|$, one finds, in the limit $\lambda \to 0$,

$$\frac{\hat{\mathcal{N}}(q = 0,\ x\lambda/|\ln \lambda|,\ \lambda)}{\hat{\mathcal{N}}(q = 0, 0, \lambda)} \approx e^{-x} . \tag{11}$$

Having noted that e^{-x} is the Laplace transform of $\delta(u - 1)$, we conclude that when $S \to \infty$, the variable $K/\sqrt{S} \ln S$ tends to unity with probability one. One can also study how the fluctuations behave for large S. Setting $K = \sqrt{S} \ln S + y\sqrt{S}$, one finds that the Laplace transform $\hat{P}(z)$ of the distribution $P(y)$ of the random variable y reads, for $S \to \infty$:

$$\hat{P}(z) = \int dy e^{-zy} P(y) = \exp[-z + (1 + z)\ln(1 + z)] , \tag{12}$$

which shows that the distribution of y is non-Gaussian, even in the limit $S \to \infty$. For example, the skewness of $P(y)$ is found to be equal to -1. In summary, we find that the average number of 'sub-entities' is equal to $\sqrt{S} \ln S$, with relative (non-Gaussian) fluctuations which go to zero as $1/\ln S$. The average size of a sub-piece is clearly equal to $\sqrt{S}/\ln S$.

Therefore, the most probable partition of a large integer S is to break it in $\sqrt{S}$ parts of size $\sqrt{S}$ (neglecting logarithms).[2] In fact, as we now show, this is not really correct. A better description is to say that one has $\sqrt{S}$ pieces of size $1, \sqrt{S}/2$ pieces of size $2, \ldots$ and one piece of size $\sqrt{S}$. More precisely, what is the average number of occurrences $N(s|S)$ of a piece of size s, given the total size S? A little reflection tells us that this is given by

$$N(s|S) \equiv \mathcal{N}(S)\langle O(s|S)\rangle = \sum_{k=1}^{\infty} \mathcal{N}(S - ks) \equiv \sum_{k=1}^{\infty} kQ(S - ks)\,, \tag{13}$$

where Q is the probability of occurrence that the number s appears exactly k times in the partition, defined by the above equation. For $1 \ll s \ll S$, $N(s|S)$ can be approximated by

$$N(s|S) \approx \mathcal{N}(S) \sum_{k=1}^{\infty} \exp\left(-\frac{bks}{2\sqrt{S}}\right) \sim \frac{\mathcal{N}(S)}{\exp(bs/2\sqrt{S}) - 1}\,. \tag{14}$$

One therefore finds the following interesting result: the size distribution of sub-sectors follows, in Sutton's model, a Bose–Einstein distribution. This distribution behaves as a power law $1/s$, for $s \ll \sqrt{S}$ and decays exponentially fast for $s \gg \sqrt{S}$. This is a directly testable prediction of Sutton's model. One can furthermore check directly that

$$\frac{\sum_{s=1}^{S} sN(s|S)}{\mathcal{N}(S)} \approx \frac{4S}{b^2} \int_0^{\infty} \mathrm{d}u\, \frac{u}{\mathrm{e}^u - 1} = S\,, \tag{15}$$

as it should.

As noticed by Sutton, the quantity $N(s|S)$ is interesting because it allows us to compute the variance $\sigma_R^2(S)$ of the absolute growth rate R, defined as

$$\sigma_R^2(S) = \overline{\langle R^2|S\rangle}\,, \tag{16}$$

where the brackets means an average over the random growth rates η_i and the overline is an average over all partitions of S. Using the fact that the η_i's are independent and of variance equal to σ_0^2, one has

$$\overline{\langle R^2|S\rangle} = \sigma_0^2 \sum_s s^2 \frac{N(s|S)}{\mathcal{N}(S)} \approx \sigma_0^2 \sum_s \frac{s^2}{\exp(bs/2\sqrt{S}) - 1}$$

$$= \frac{2^{5/2}3^{3/2}\zeta(3)}{\pi^3}\, \sigma_0^2 S^{3/2} = 1.13955....\sigma_0^2 S^{3/2}\,. \tag{17}$$

This is Sutton's result: the conditional variance of the absolute return grows as $S^{3/2}$, therefore the variance of the *relative* return $r = R/S$ decays as $S^{-1/2}$, which is equivalent

[2] The $\sqrt{S}$ scaling found here for the typical size of sub-pieces allows to shed light on a totally unrelated problem, that of 'branched polymers' in high dimensions [6,7]. In the absence of steric constraints, the end-to-end distance of branched polymers grows like $N^{1/4}$, where N is the total number of monomers. The scaling comes from the fact that a typical linear strand of the polymer contains $\sim N^{1/2}$ monomers, each of which behaving as a random walk in space. The analysis given here suggests that the number of independent linear strands (that plays the role of K) scales as $N^{1/2} \ln N$, with relative fluctuations that tend to zero, and that the size distribution of these strands should decay as $1/n$ for $n \ll N^{1/2}$.

to the statement that $\beta=\frac{1}{4}$ [4]. In intuitive terms, the total absolute return is the random sum of $\sim\sqrt{S}$ different terms, all of order $\sim\sqrt{S}$, which gives a random number of order $\sqrt{S}\sqrt{\sqrt{S}}\sim S^{3/4}$. This rough (and slightly incorrect) argument actually suggests that the absolute growth rate is the sum of a large number ($\sqrt{S}$) of random variables, and therefore should be Gaussian for large S. We can show this more precisely by computing the kurtosis κ of R, defined as

$$\kappa=\frac{\overline{\langle R^4|S\rangle}}{\overline{\langle R^2|S\rangle}^2}-3\,. \tag{18}$$

We assume that the individual growth rates η have a finite kurtosis given by κ_0. Therefore,

$$\overline{\langle R^4|S\rangle}=\sigma_0^4\left[(3+\kappa_0)\overline{\sum_i s_i^4}+3\overline{\sum_{i\neq j}s_i^2s_j^2}\right]=\sigma_0^4\left[\kappa_0\overline{\sum_i s_i^4}+3\overline{\sum_{i,j}s_i^2s_j^2}\right]. \tag{19}$$

The first term is easy to compute using $N(s|S)$ and one finds

$$\overline{\sum_i s_i^4}=\sum_s s^4\frac{N(s|S)}{\mathcal{N}(S)}\approx\frac{2^{11/2}3^{7/2}}{\pi^5}\zeta(5)S^{5/2}=7.17114....S^{5/2}\,. \tag{20}$$

The second term is more subtle since one needs to know the correlation of the number of occurrences of two integers s,s' involved in the partition of S, $\langle O(s|S)O(s'|S)\rangle$. This quantity can be obtained similarly to $N(s|S)$. For $s=s'$, one has

$$\mathcal{N}(S)\langle O(s|S)^2\rangle\equiv\sum_{k=1}^{\infty}k^2Q(S-ks)=\sum_{k=1}^{\infty}(2k-1)\mathcal{N}(S-ks)\,, \tag{21}$$

which can be again be approximated as

$$\langle O(s|S)^2\rangle\approx\sum_{k=1}^{\infty}(2k-1)\exp\left(-\frac{bks}{2\sqrt{S}}\right)=\frac{\exp(bs/2\sqrt{S})+1}{(\exp(bs/2\sqrt{S})-1)^2}\,. \tag{22}$$

Therefore,

$$\sum_{s=1}^{S}s^4[\langle O(s|S)^2\rangle-\langle O(s|S)\rangle^2]\approx\left(\frac{2\sqrt{S}}{b}\right)^5\int_0^{\infty}\mathrm{d}u u^4\frac{\mathrm{e}^u}{(\mathrm{e}^u-1)^2}$$
$$=\frac{2^{9/2}3^{3/2}}{5\pi}S^{5/2}=7.48509....S^{5/2}\,. \tag{23}$$

The terms with $s\neq s'$ can be computed from

$$\langle O(s|S)O(s'|S)\rangle\approx\sum_{k=1,\ell}^{\infty}\exp\left(-\frac{bks}{2\sqrt{S}}-\frac{b\ell s'}{2\sqrt{S}}-\frac{b}{8S^{3/2}}(ks+\ell s')^2+\cdots\right)\,. \tag{24}$$

The reason we took one extra term in the above expansion is that $\langle O(s|S)O(s'|S)\rangle - \langle O(s|S)\rangle\langle O(s'|S)\rangle$ is zero to first order. The non-zero correlation comes from the term $k\ell ss'$ in the above expression. To lowest order, one finally finds

$$\sum_{s\neq s'=1}^{S} s^2 s'^2 [\langle O(s|S)O(s'|S)\rangle - \langle O(s|S)\rangle\langle O(s'|S)\rangle]$$

$$\approx -\frac{2^{19/2}3^{5/2}}{\pi^5}[\zeta(3)]^2 S^{3/2} = -53.2953\ldots S^{3/2}\,. \tag{25}$$

This contribution is a factor S smaller than the other two contributions, but has a rather large prefactor. To leading order in S, the final result reads

$$\kappa \approx \frac{1}{\sqrt{S}}[5.52232\ldots\kappa_0 + 5.76408\ldots]\,. \tag{26}$$

The conclusion is that even if the growth rates of the sub-sectors are non-Gaussian, the kurtosis of the aggregate growth rate decreases as $1/\sqrt{S}$ for large S. This is expected on general grounds, since we have seen above that the number of independent sub-entities is of order $\sqrt{S}$, and the kurtosis of a sum decreases as the inverse of the number of independent terms in a sum. The second contribution comes from the fluctuations of the numbers of terms in the sum.

Therefore, *asymptotically*, the rescaled aggregate growth rate $rS^{1/4}$ is found to be Gaussian in Sutton's model. However, for finite S, there are important corrections to this asymptotic result: suppose that the initial kurtosis of η is equal to 3, which is the case when η is distributed according to a symmetric exponential. Take a reasonable value $S = 100$. Then the residual kurtosis of the growth rate is still quite large, ~ 2.2. Hence, significant deviations from a Gaussian distribution may be observed in reality, but should *diminish* as S becomes large, at variance with empirical results. We shall come back to this issue in Section 4.

3. An alternative model

3.1. Definition of the model

We now discuss another model, where we assume that companies are formed by aggregating entities that have a certain a priori distribution of sizes, that we choose to be a power-law in its tail. The motivation for this is two-fold. First, the distribution of company sizes in a country is known to follow a Pareto (power-law) distribution. Since the scaling law for the variance of the growth rate also seems to hold at the country level, one could indeed argue that the actual distribution of company sizes should play a role. Second, there is a quite general and plausible *dynamical* model that leads to a power-law distribution of sizes. Assume, as in Sutton's model, that each sub-entity in a company has a random growth rate. The role of the business management is, to a certain extent, to redistribute the income of each sector of activity such as to help the less performing ones to catch up. Therefore, a reasonable dynamical model for the

size $s_i(t)$ of a given sub-entity is

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = \gamma\left(\frac{1}{K}\sum_{j=1}^{K} s_j(t) - s_i(t)\right) + \eta_i(t)s_i(t)\,, \tag{27}$$

where the first two terms describe redistribution of resources among the sub-entities, and the last term the random growth rate. The parameter γ measures the strength of the redistribution policy. It can be shown that the stationary distribution for such a stochastic process has a power-law tail, $p(s) \sim s^{-1-\mu}$, with $\mu = 1 + \gamma/\sigma_0^2$. (See the detailed discussion and generalization in Ref. [8], and also Refs. [9,10] for alternative models.) The above 'redistribution' model is only meant to be an illustration of how Pareto tails could emerge dynamically, but is not logically related to the following discussion, where we only *assume* that the a priori distribution of the size of sub-entities has a power-law tail:

$$p(s) \approx \frac{\mu s_0^{\mu}}{s^{1+\mu}} \quad (s \to \infty)\,. \tag{28}$$

We also assume that a company is composed of an arbitrary number K of such sub-entities, with a certain a priori weight $\mathcal{Q}(K)$. This means that if one chooses randomly a company in a country, there is a probability proportional to $\mathcal{Q}(K)$ for this company to contain exactly K sectors. We will see below that $\mathcal{Q}(K)$ can be inferred from empirical data. The unnormalized distribution of growth rates for a given company size S reads, in this new model:

$$\mathcal{N}(R,S) = \sum_{K=1}^{\infty} \mathcal{Q}(K) \int \prod_{i=1}^{K} p(s_i)\,\mathrm{d}s_i\,\delta\left(S - \sum_{i=1}^{K} s_i\right) \int \prod_{i=1}^{K} P(\eta_i)\,\mathrm{d}\eta_i \times \delta\left(R - \sum_{i=1}^{K} s_i\eta_i\right)\,. \tag{29}$$

3.2. Distribution of company sizes

Let us first establish some results on the size distribution of companies $\mathcal{N}(S) = \int \mathrm{d}R\,\mathcal{N}(R,S)$, a quantity much studied in a different context in Ref. [11]. This will also enable us to relate $\mathcal{Q}(K)$ to this empirically observable quantity. We first study the case the simplest case $\mathcal{Q}(K) = 1$. The following results are obtained using Laplace transforms, as above. We write

$$\hat{\mathcal{N}}(q,\lambda) = \int_0^{\infty} \mathrm{d}S \int_{-\infty}^{\infty} \mathrm{d}R\,\exp[\mathrm{i}qR - \lambda S]\,\mathcal{N}(R,S) = \sum_{K=1}^{\infty}\left[\int \mathrm{d}s\,\mathrm{d}\eta\,p(s)P(\eta)\mathrm{e}^{\mathrm{i}qs\eta - \lambda s}\right]^K\,. \tag{30}$$

For simplicity, we assume that $P(\eta)$ is Gaussian with unit variance, and introduce the quantity $g(q,\lambda)$ as

$$g(q,\lambda) = \int \mathrm{d}s\, p(s)[1 - \mathrm{e}^{-q^2 s^2/2 - \lambda s}] \,, \tag{31}$$

in terms of which one finally has

$$\hat{\mathcal{N}}(q,\lambda) = \frac{1}{g(q,\lambda)} \,. \tag{32}$$

All the following asymptotic results will only depend on the behaviour of $g(q,\lambda)$ in the limit $q,\lambda \to 0$. If we first study the case $q=0$ from which $\mathcal{N}(S)$ is deduced, one finds that one has to distinguish the cases $\mu < 1$ and $\mu > 1$ [11]. The small λ behaviour of g is found to be

$$g(q=0,\lambda) \approx \Gamma(1-\mu)(s_0\lambda)^{\mu} \quad (\mu<1), \quad g(q=0,\lambda) \approx \lambda\langle s\rangle \quad (\mu>1) \,. \tag{33}$$

For $\mu > 1$, the average size of a sub-entity is finite and equal to $\langle s\rangle$. Inverting the Laplace transform then leads to

$$\mathcal{N}(S) \approx \frac{1}{\langle s\rangle} \,, \tag{34}$$

for $\mu > 1$, whereas for $\mu < 1$, one has

$$\mathcal{N}(S) \approx \frac{\sin \pi\mu}{\pi} \frac{1}{s_0} \left(\frac{s_0}{S}\right)^{1-\mu} \,. \tag{35}$$

The case $\mu=1$ is special and involves logarithmic corrections. Intuitively, the difference of behaviour comes from the fact that when $\mu > 1$, the typical number of sub-entities behaves as $K \sim S/\langle s\rangle$, whereas when $\mu < 1$, a single sub-entity represents a sizeable fraction of the whole company and $K \sim S^{\mu} \ll S$.

Assuming now that $\mathcal{Q}(K)$ decays as a power-law $\mathcal{Q}(K) \sim K^{-1-\alpha}$ and that $\mu > 1$, we find, using the same method, the following result for $P(S)$:

$$\mathcal{N}(S) \sim \frac{1}{S^{1+\alpha}} \quad (\alpha \leqslant \mu), \quad \mathcal{N}(S) \sim \frac{1}{S^{1+\mu}} \quad (\alpha \geqslant \mu) \,. \tag{36}$$

The case $\alpha \geqslant \mu$ corresponds to a situation where large companies only contain a small number of sectors (see below). This is not very plausible; furthermore, this would lead to a variance of the growth rate R that grows proportionally to size S, i.e., $\beta = 0$, which is not compatible with empirical data. Therefore, we will assume in the following $\alpha \leqslant \mu$. In this case, there is a direct relation between the tail, of $\mathcal{Q}(K)$ and the tail of the size distribution of companies. Empirically, α is found to be close to unity: $\alpha \approx 1.05$ [12].

3.3. Fluctuations of the growth rate

We now turn to the prediction of this model for the growth rate fluctuations. One need to consider three cases: $\mu > 2, 1 < \mu < 2$, $\mu < 1$. The case $1 < \mu < 2$ is, as we will show, the interesting one. In the relevant situation where $\alpha \leqslant \mu$, one can show that the value of β and the shape of the rescaled distributions are independent of the value of α, and we choose in the following, for simplicity $\mathcal{Q}(K)$ to be constant.

We now need to study $\hat{\mathcal{N}}(q,\lambda)$ with $q \neq 0$, that gives access to the distribution of the growth rate. This case can be treated by identifying the correct scaling region in the q,λ plane, which means, in concrete terms, the scaling relation between R and S. For example, when $\mu > 2$, one expects the Central Limit Theorem to hold, suggesting $R \sim \sqrt{S}$. So we set $q = \theta\sqrt{\lambda}$, and take the limit $\lambda \to 0$. If $\mu > 2$, one finds

$$g(q,\lambda) \approx \lambda \left[\langle s \rangle + \frac{\theta^2}{2} \langle s^2 \rangle \right] + \cdots , \tag{37}$$

where the $\cdots$ refers to higher-order terms in λ, the precise form of which depend on the value of μ. Therefore, in this regime,

$$\hat{\mathcal{N}}(q,\lambda) \approx \frac{1}{\lambda[\langle s \rangle + (\theta^2/2)\langle s^2 \rangle]} . \tag{38}$$

Now, we introduce the probability $P(R|S)$ to observe a certain growth R given S. Then, by definition,

$$\hat{\mathcal{N}}(q,\lambda) = \int_0^\infty \mathrm{d}S \int_{-\infty}^\infty \mathrm{d}R \exp[\mathrm{i}qR - \lambda S] \mathcal{N}(S) P(R|S) . \tag{39}$$

Assuming that $P(R|S) = S^{-1/2}\Pi(RS^{-1/2})$ and using the above result for $\mathcal{N}(S)$ leads to

$$\hat{\mathcal{N}}\left(q = \theta\sqrt{\lambda}, \lambda\right) \approx \frac{1}{\lambda\langle s \rangle} \int_0^\infty \mathrm{d}u \int_{-\infty}^\infty \mathrm{d}v \exp\left[\mathrm{i}\theta v\sqrt{u} - u\right] \Pi(v) , \tag{40}$$

where we have set $\lambda S = u$ and $R = v\sqrt{S}$. It is now easy to see that Eqs. (38) and (40) are satisfied if

$$\Pi(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right), \quad \sigma^2 = \frac{\langle s^2 \rangle}{\langle s \rangle} . \tag{41}$$

Therefore, in the case $\mu > 2$, the variance of the relative growth rate decreases as $S^{-1/2}$ (i.e., $\beta = \frac{1}{2}$), and the distribution of growth rates is Gaussian.

More interesting is the case $1 < \mu < 2$. It turns out that in this regime, the correct scaling is $P(R|S) = S^{-1/\mu}\Pi(RS^{-1/\mu})$ and $q = \theta\lambda^{1/\mu}$. In this regime, one now has, for small λ

$$g(q,\lambda) \approx \lambda[\langle s \rangle + \mu|s_0\theta|^\mu I(\mu)] + \cdots , \tag{42}$$

where

$$I(\mu) = \int_0^\infty \frac{\mathrm{d}t}{t^{1+\mu}} (1 - \mathrm{e}^{-t^2/2}) = -2^{-1-\mu/2}\Gamma\left(-\frac{\mu}{2}\right) . \tag{43}$$

Now, it is easy to show that the scaling function $\Pi(v)$ is now precisely a symmetric Levy stable distribution of index $\mu, L_\mu(v)$. This comes from the fact that the Fourier transform of $L_\mu(v)$ gives $\exp(-Au|\theta|^\mu)$, where A is a constant, so that the integral over u in Eq. (40) now reproduces Eq. (42).

However, this is not the whole story. The reason is that a direct computation of the variance of R (from the derivative of $g(q,\lambda)$ with respect to q^2 at $q = 0$) leads an apparently contradictory scaling, since

$$\langle R^2|S \rangle \propto S^{3-\mu} \tag{44}$$

instead of $S^{2/\mu}$ as one might have naively expected from the scaling form of $P(R|S)$. One should now remember that Lévy stable distributions $L_\mu(v)$ with $\mu < 2$ have tails decaying as $v^{-1-\mu}$, and thus a formally infinite variance. This means that $\langle R^2|S\rangle$ is actually dominated by the region where R is of order S, such that indeed:

$$\langle R^2|S\rangle \approx S \int_{-S}^{+S} \frac{R^2\,\mathrm{d}R}{|R|^{1+\mu}} \sim S^{3-\mu}\,. \tag{45}$$

Therefore, the Lévy stable distribution only holds in the scaling region $R \sim S^{1/\mu}$. For $R \sim S \gg S^{1/\mu}$ the distribution is truncated. When $\mu \to 1$, the truncation 'invades' the scaling regime, and the result becomes again different for $\mu < 1$, see below.

The conclusion of this analysis is that the variance of the relative growth rate $r = R/S$ scales in this regime with an exponent $\beta = (\mu - 1)/2$, that interpolates between the standard value $\beta = \frac{1}{2}$ for $\mu = 2$ and $\beta = 0$ for $\mu = 1$ (although these marginal cases are affected by logarithmic corrections). However, the surprising result is that in this regime the distribution of R does *not* re-scale as a function of rS^β but rather as $rS^{(\mu-1)/\mu}$.[3] We will discuss this in relation with empirical results in the next section.

When $\mu < 1$, it is easy to show that now $R \sim S$, i.e., $\beta = 0$, which disagrees with empirical results. Furthermore, the result one finds for the scaling function Π is no longer universal. When $\mu > 1$, the scaling function was universal in the sense that its shape only relied on the finiteness of the variance of η. When $\mu < 1$, on the other hand, only a finite number of terms (sub-entities) contribute to the sum R, and one cannot expect a Central Limit Theorem to hold. When $P(\eta)$ is Gaussian of variance σ_0^2, all moments of $P(R|S)$ can be computed using the method of Ref. [15]. One finds for example

$$\langle R^2|S\rangle = \frac{1-\mu}{1+\mu}\,\sigma_0^2 S^2 \tag{46}$$

and

$$\langle R^4|S\rangle = 3\left[\frac{(3-\mu)(2-\mu)(1-\mu) + 2\mu(1-\mu)^2}{(3+\mu)(2+\mu)(1+\mu)}\right]\sigma_0^4 S^4\,. \tag{47}$$

Numerically, we have found that $P(R|S)$ could be rather well fitted by a 'stretched Gaussian' form, $\exp -(R/S)^\alpha$ with $\alpha < 2$. For example, for $\mu = \frac{1}{2}$, we found $\alpha \approx \frac{4}{3}$. This cannot be exact, however, since the exact kurtosis is found to be 1.37143, whereas the kurtosis of the stretched Gaussian with $\alpha = \frac{4}{3}$ is 1.22219. Note that $\langle R^4|S\rangle$ would have a different value if $P(\eta)$ was non-Gaussian: this shows that for $\mu < 1$ the distribution $\Pi(v)$ is non-universal.

3.4. Conditional distribution of sector sizes

Finally, one can also compute in this model the conditional distribution of sector sizes, $P(s|S)$, which depends on the value of α. When $\alpha \leqslant \mu$, we find that $P(s|S)$ is the sum of two contributions: one power-law regime $s^{-1-\mu}$ for $s \ll S$ which reflects

[3] For other situations where this 'anomalous scaling' occur, see Refs. [13,14].

the a priori distribution of sector sizes, and a small 'hump' for $s \sim S$ of height which vanishes for large S:

$$P(s|S) \approx \frac{\mu s_0^{\mu}}{s^{1+\mu}} \quad (s \ll S), \qquad P(s|S) \approx \frac{F(s/S)}{S^{1+\mu-\alpha}} \quad (s \sim S)\,, \tag{48}$$

where $F(.)$ is a certain scaling function of order unity, that vanishes for $s > S$. For $\mu < \alpha$, on the other hand, the hump survives when $S \to \infty$ whereas the power-law regime disappears. In other words, when α is larger than μ, the typical number of sectors K^* tends to be small and the typical size of the sectors is of the order of S itself.

3.5. Stability upon aggregation

As mentioned in the introduction, the scaling of GNP growth rates is empirically found to be very similar to the scaling of company growth [3]. In this respect, it is worth noting that Sutton's construction is not stable upon aggregation: aggregating companies characterized by an exponent $\beta = \frac{1}{4}$ using Sutton's prescription at the country level, leads to an exponent $\beta = \frac{3}{8}$. In our model, on the other hand, stability upon aggregation is by construction satisfied. The argument is very simple, and relies on the fact that the results are independent of the value of the company size exponent α, provided $\alpha < \mu$. The idea is to consider the GNP itself as the sum of independent sectors, i.e., to remove the 'shells' that define companies, which are an intermediate level of clustering. A country is therefore in this description a 'super-company' with many sectors. The sectors are the same than previously, so they have the very same Pareto tail of exponent μ for their size distribution. Now we just have to assume that there is a given distribution $Q'(K)$ that describes the distribution of the number of independent sectors in different countries. If $Q'(K)$ has a Pareto tail with exponent α' with $\alpha' < \mu$, we can repeat the above arguments and find the same value for the exponent $\beta = (\mu - 1)/2$ at the country level.

4. Discussion—comparison with empirical data

We have shown how several interesting asymptotic predictions of Sutton's model could be derived. Apart from Sutton's central result, namely that the root mean square of the growth rate decreases with the company size S as $S^{-1/4}$ (i.e., $\beta = \frac{1}{4}$), we have shown that the distribution of growth rate should be asymptotically Gaussian, with a kurtosis that decays as $S^{-1/2}$. The first result is, as noticed by Sutton, in rather good agreement with the empirical results of Stanley et al. [1], although the value of the exponent β is slightly smaller. The second result is however problematic, since in this model one should find a rescaled distribution of growth rates that progressively deforms with S as to become Gaussian for very large S, whereas the data indicates that the rescaled distribution is actually to a good approximation independent of S *and* non-Gaussian. A closer look at the data of Stanley et al. in fact suggests that non-Gaussian tails are *more pronounced* for larger companies [2].

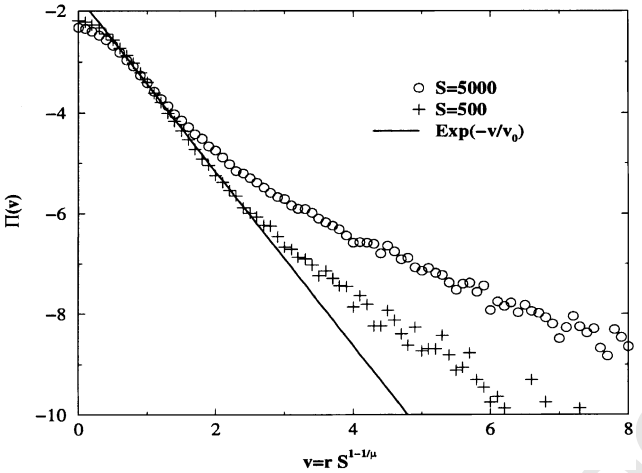

Fig. 1. Distribution $\Pi(v)$ of the rescaled returns for $v > 0$ (in a linear-log representation) and for two values of S: $S \approx 500$ and 5000, and for $\mu = 1.3$. A simple exponential form, as suggested in Ref. [1], is shown for comparison. Note that for this range of S, the distributions do not re-scale, except in the 'central' region. For large values of S, the distribution should converge towards a Lévy distribution of index $\mu = 1.3$: one can clearly see the tails getting fatter as S increases, as is also the case for empirical data.

We have then explored an alternative to Sutton's model, where the size of the 'sub-entities' is postulated to be asymptotically a power-law (Pareto) with an unknown exponent μ, that we relate to β—we thus abandon the idea of directly *predicting* β. Our model is motivated by the ubiquitous observation of Pareto distributions for company sizes, and by a simple dynamical model that indeed leads to a stationary power-law distribution of sizes. In this respect, it is not obvious how one would write a natural dynamics for sector growth that leads to Sutton's 'microcanonical' ensemble where all partitions are equiprobable. As a function of μ, we have found three qualitatively different regimes. In particular, when $1 \leqslant \mu \leqslant 2$, we find that $\beta = (\mu - 1)/2$. The empirical value $\beta = 0.15$ corresponds to $\mu = 1.30$, which is indeed larger than the value of $\alpha \approx 1.05$ reported for firm sizes in Ref. [12], as required for the consistency of our analysis. Our model then predicts an S independent distribution for the growth rate multiplied by $S^{(\mu-1)/\mu}$ (and not by S^{β}), which is a symmetric Lévy stable distribution. Note however that these are asymptotic results that require $S^{(\mu-1)/\mu} \gg 1$, such that the scaling region is not affected by truncation effects (see the discussion after Eq. (44)). For finite S and μ close to one, one expects strong finite size effects, and a very slow convergence towards the asymptotic value. This is why numerical simulations are needed to explore the moderate S regime. We show in Fig. 1 the distribution of rescaled returns obtained from a numerical simulation for $\mu = 1.30$, a Gaussian $P(\eta)$ and for $S \sim 500$ and $S = 5000$. Notice $\Pi(v)$ can be very roughly approximated by a symmetric exponential for small enough S: $\Pi(v) = \exp(-|v|/v_0)$, as suggested by the empirical data—at least in a restricted range of v. The systematic deviations from this form both at small values of v and at large v are qualitatively similar to the ones observed on the

data (see Ref. [2]): the empirical $\Pi(v)$ is actually parabolic for small values of v and decays slower than exponentially at large v. One also observes finite size effects on the empirical data [2]: as S increases, the tail of the distribution becomes fatter and fatter. This is expected in our model since asymptotically $\Pi(v)$ should converge to a Lévy distribution with a power-law tail, which is indeed fatter than an exponential. Note that we expect only qualitative agreement with empirical data, since the assumption that $P(\eta)$ is Gaussian at the sector level is probably incorrect and does influence the detailed shape of $\Pi(v)$ for finite S.

It would be extremely interesting to obtain direct empirical information on the conditional distribution of the size s and total number K of the sub-entities for a fixed S. We have seen that Sutton's model predicts a Bose–Einstein distribution for s, that behaves as $1/s$ for $s \ll \sqrt{S}$, and beyond which it falls rapidly, whereas K becomes peaked around the value $\sqrt{S}\ln S$. In our model, on the other hand, the conditional distribution of s is, as soon as $\mu \geqslant \alpha$ and for $s \ll S$, identical to the a priori distribution $p(s) \sim s^{-1-\mu}$, and the total number K peaks around the value $S/\langle s \rangle$. Therefore, a tangible difference between the two models is that the power-law regime has an exponent 1 in the Sutton model and the size of the sectors rarely exceeds $\sqrt{S}$, whereas the distribution is a power-law with exponent $1+\mu \approx 2.35$ up to S in our model (with possibly a small hump for $s \sim S$, see Eq. (48)). We hope that these falsifiable predictions of the two descriptions, as well as the quantitative description of the rescaled distribution of growth rates given above, will motivate further empirical and theoretical research, and help elucidate the 'scaling puzzle' of company growth.

Note added: While completing this work, X. Gabaix sent us a very interesting preprint where related arguments (although in details quite different from ours) are discussed. See: X. Gabaix, *Power-laws and the origin of the business cycle*, working paper, October 2002.

Acknowledgements

We want to thank X. Gabaix, M. Potters, J. Scheinkman and in particular J. Sutton for interesting discussions. J.P.B. also wants to thank the organizers of the Bali conference on Econophysics that took place in August 2002 and that motivated this work.

References

[1] M.H.R. Stanley, L.A.N. Amaral, S. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, Nature 319 (1996) 804.

[2] L.A.N. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, Scaling behavior in economics: I. Empirical results for company growth, J. Phys. I France 7 (1997) 621;
S.V. Buldyrev, L.A.N. Amaral, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, Scaling behavior in economics: II. Modeling of company growth, J. Phys. I France 7 (1997) 635.

[3] Y. Lee, L.A.N. Amaral, D. Canning, M. Meyer, H.E. Stanley, Phys. Rev. Lett. 81 (1998) 3275;
D. Canning, L.A.N. Amaral, Y. Lee, M. Meyer, H.E. Stanley, Econom. Lett. 60 (1998) 335.

[4] J. Sutton, Physica A 312 (2002) 577.
[5] G. Andrews, The Theory of Partitions, Cambridge University Press, Cambridge, 1998.
[6] P.G. de Gennes, Biopolymers 6 (1968) 715.
[7] T. Lubensky, J. Isaacson, Phys. Rev. A 20 (1979) 2130.
[8] J.-P. Bouchaud, M. Mézard, Physica A 282 (2000) 536.
[9] X. Gabaix, Quart. J. Econom. 114 (1999) 739.
[10] J.-P. Bouchaud, Quant. Finance 1 (2000) 105.
[11] F. Bardou, J.P. Bouchaud, A. Aspect, C. Cohen-Tannoudji, Lévy statistics and Laser cooling, Cambridge University Press, Cambridge, 2002.
[12] ■
[13] ■
[14] ■
[15] ■